

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:54:57 ; Search time 73.3433 Seconds
(without alignments)
69.343 Million cell updates/sec

Title: US-09-641-801-3
Perfect score: 98
Sequence: 1 DQPPDVEKPDLPFQVQS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	98	100.0	18	4	AAB72502	Aab72502 Colostrin
2	98	100.0	18	4	AAB59325	Aab59325 Ewe colos
3	98	100.0	18	4	AAB72248	Aab72248 Colostrin
4	98	100.0	18	4	AAB72534	Aab72534 Colostrin
5	98	100.0	18	5	AAO14579	Aao14579 Neural ce
6	98	100.0	18	5	AAM51038	Aam51038 Colostrin
7	98	100.0	18	5	AAE20230	Aae20230 Colostrin
8	98	100.0	19	4	AAB59355	Aab59355 Ewe colos
9	51	52.0	1047	6	ABU28927	Abu28927 Protein e

10	48	49.0	377	4	AAM79068	Aam79068	Human pro
11	48	49.0	377	5	ABB83472	Abb83472	Human cyt
12	48	49.0	377	7	ADE47756	Ade47756	Human NOV
13	48	49.0	541	4	AAU35153	Aau35153	Enterococ
14	47	48.0	516	5	ABB91350	Abb91350	Herbicida
15	47	48.0	719	5	ABB91351	Abb91351	Herbicida
16	46.5	47.4	120	3	AAG56012	Aag56012	Arabidops
17	46.5	47.4	126	3	AAG56011	Aag56011	Arabidops
18	46.5	47.4	142	3	AAG56010	Aag56010	Arabidops
19	46	46.9	176	4	AAU55015	Aau55015	Propionib
20	46	46.9	176	6	ABM51534	Abm51534	Propionib
21	46	46.9	176	6	ABM64933	Abm64933	Propionib
22	46	46.9	180	3	AAY45073	Aay45073	Rat amelo
23	46	46.9	194	3	AAY45072	Aay45072	Rat amelo
24	45	45.9	26	3	AAB20811	Aab20811	PEST sequ
25	45	45.9	181	4	ABG29328	Abg29328	Novel hum
26	45	45.9	188	4	AAB63292	Aab63292	Human bre
27	45	45.9	188	4	AAB63288	Aab63288	Human bre
28	45	45.9	215	4	AAB63316	Aab63316	Human bre
29	45	45.9	233	4	AAB63317	Aab63317	Human bre
30	45	45.9	383	4	ABG27213	Abg27213	Novel hum
31	45	45.9	383	4	ABG09426	Abg09426	Novel hum
32	45	45.9	848	3	AAB58428	Aab58428	Lung canc
33	45	45.9	858	6	ABR64269	Abr64269	Angiogene
34	45	45.9	858	7	ADE54205	Ade54205	Human Pro
35	45	45.9	858	7	ADE54209	Ade54209	Human Pro
36	45	45.9	859	5	ABB57101	Abb57101	Mouse isc
37	45	45.9	859	7	ADE54207	Ade54207	Rat Prote
38	45	45.9	859	7	ADE54203	Ade54203	Rat Prote
39	45	45.9	872	2	AAY07110	Aay07110	Colon can
40	45	45.9	1572	5	ABP65360	Abp65360	Bifidobac
41	44.5	45.4	852	6	ABU09534	Abu09534	Mouse DNA
42	44.5	45.4	1520	4	ABB64314	Abb64314	Drosophil
43	44	44.9	167	6	ABP72774	Abp72774	Human tra
44	44	44.9	192	4	AAB94225	Aab94225	Human pro
45	44	44.9	192	6	ABU00368	Abu00368	Human nov

ALIGNMENTS

RESULT 1

AAB72502

ID AAB72502 standard; peptide; 18 AA.

XX

AC AAB72502;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #3.

XX

KW Dermatological; oxidative stress regulator; colostrinin.

XX

OS Unidentified.

XX

PN WO200112650-A2.

XX

PD 22-FEB-2001.
 XX
 PF 17-AUG-2000; 2000WO-US022665.
 XX
 PR 17-AUG-1999; 99US-0149310P.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Stanton GJ, Hughes TK, Boldogh I;
 XX
 DR WPI; 2001-218342/22.
 XX
 PT Modulating oxidative stress level in a cell, involves contacting the cell
 PT with an oxidative stress regulator selected from colostrinin, its
 PT constituent peptide, analog or their combinations.
 XX
 PS Claim 6; Page 25; 48pp; English.
 XX
 CC The present invention relates to a method for modulating the oxidative
 CC stress level in a cell or a patient, comprising contacting the cell with,
 CC or administering to the patient, an oxidative stress regulator selected
 CC from colostrinin, or its constituent peptide (e.g. the present peptide),
 CC to change the level of an oxidising species in the cell. The method can
 CC be used to treat oxidative damage to skin, by decreasing or preventing an
 CC increase in the level of damage to a biomolecule of the patient
 XX
 SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 4; Length 18;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DQPPDVEKPDLPFQVQS 18
 |||||
 Db 1 DQPPDVEKPDLPFQVQS 18

RESULT 2

AAB59325

ID AAB59325 standard; peptide; 18 AA.

XX

AC AAB59325;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment B-10.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;

KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX
 PR 02-JUN-1999; 99GB-00012852.
 XX
 PA (REGE-) REGEN THERAPEUTICS PLC.
 XX
 PI Georgiades JA;
 XX
 DR WPI; 2001-071058/08.
 XX
 PT Peptides having an N-terminal amino acid sequence isolated from
 PT colostrinin for treating e.g. disorders of the central nervous system and
 PT immune system, viral and bacterial infections, and diseases characterized
 PT by amyloid plaques.
 XX
 PS Claim 7; Page 27; 63pp; English.
 XX
 CC The present invention provides the sequences of a number of peptides
 CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide
 CC fragment of colostrum. These peptides can be used in the treatment of
 CC central nervous system disorders such as senile dementia, Parkinson's
 CC disease, Alzheimer's disease, psychosis and neurosis, immune system
 CC disorders such as bacterial and viral infections, to improve the
 CC development of a child's immune system, as a dietary supplement, and to
 CC promote the dissolution of beta-amyloid plaques
 XX
 SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 4; Length 18;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
 |||||
 Db 1 DQPPDVEKPDLPFQVQS 18

RESULT 3

AAB72248

ID AAB72248 standard; peptide; 18 AA.

XX

AC AAB72248;

XX

DT 14-MAY-2001 (first entry)

XX

DE Colostrinin derived cytokine inducing peptide SEQ ID 3.

XX

KW Colostrinin; immune response; cytokine; blood cell proliferation;
 KW central nervous system disorder; neurological disorder; mental disorder;
 KW dementia; neurodegenerative disease; Alzheimer's disease; psychosis;
 KW neurosis; infection.

XX

OS Synthetic.

XX

PN WO200111937-A2.

XX

PD 22-FEB-2001.

XX

PF 17-AUG-2000; 2000WO-US022818.
 XX
 PR 17-AUG-1999; 99US-0149311P.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 PA (REGE-) REGEN THERAPEUTICS PLC.
 XX
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;
 XX
 DR WPI; 2001-202804/20.
 XX
 PT Inducing a cytokine and modulating an immune response, useful for
 PT treating central nervous system diseases and bacterial and viral
 PT infections, comprises administering colostrinin as an immunological
 PT regulator.
 XX
 PS Claim 1; Page 34; 50pp; English.
 XX
 CC Sequences AAB72246 - AAB72275 represent peptides derived from clostrinin,
 CC a proline rich polypeptide aggregate contained in colostrum. The peptides
 CC have immune response modulatory activity, and are capable of inducing
 CC cytokines. Colostrinin and its derived peptides are useful for inducing
 CC cytokine production, for modulating an immunological response and for
 CC inducing blood cell proliferation. The peptides are useful in the
 CC treatment of disorders of the central nervous system, neurological
 CC disorders, mental disorders, dementia, neurodegenerative diseases,
 CC Alzheimer's disease, motor neurone disease, psychosis, neurosis, chronic
 CC disorders of the immune system, bacterial and viral infections and
 CC acquired immunological deficiencies
 XX
 SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 4; Length 18;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
 |||||
 Db 1 DQPPDVEKPDLPFQVQS 18

RESULT 4

AAB72534

ID AAB72534 standard; peptide; 18 AA.

XX

AC AAB72534;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #3.

XX

KW Neuroprotective; neural cell differentiation regulator; colostrinin;
 KW colostrum.

XX

OS Unidentified.

XX

PN WO200112651-A2.

XX
PD 22-FEB-2001.
XX
PF 17-AUG-2000; 2000WO-US022774.
XX
PR 17-AUG-1999; 99US-0149633P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Boldogh I;
XX
DR WPI; 2001-226545/23.
XX
PT Use of colostrinin, its constituent peptide or analog as a neural cell
PT regulator, for promoting neural cell differentiation and treating damaged
PT neural cells in a patient.
XX
PS Claim 6; Page 21; 35pp; English.
XX
CC The present invention relates to a method for promoting neural cell
CC differentiation and treating damaged neural cells, using colostrinin and
CC colostrinin constituent peptides (e.g. the present peptide) as a neural
CC cell regulator. Colostrinin is a polypeptide complex found in colostrum
XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
| | | | | | | | | | | | | | | |
Db 1 DQPPDVEKPDLPFQVQS 18

RESULT 5

AAO14579

ID AAO14579 standard; peptide; 18 AA.

XX

AC AAO14579;

XX

DT 27-MAY-2002 (first entry)

XX

DE Neural cell regulatory colostrinin peptide 3.

XX

KW Neural cell differentiation; neural cell regulator; colostrinin peptide;

KW neural cell formation; proline-rich polypeptide aggregate; colostrum;

KW neural cell treatment.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Modified-site 18

FT /note= "Optional C-terminal amide"

XX

PN WO200213851-A1.

XX

PD 21-FEB-2002.
 XX
 PF 17-AUG-2000; 2000WO-US022777.
 XX
 PR 17-AUG-2000; 2000WO-US022777.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Boldogh I, Stanton JG, Hughes TK;
 XX
 DR WPI; 2002-269152/31.
 XX
 PT Promoting cell differentiation in a patient involves use of blood cell
 PT regulator selected from colostrinin, its constituent peptide and/or
 PT analog.
 XX
 PS Claim 7; Page 21; 37pp; English.
 XX
 CC The invention comprises a method for promoting cell differentiation (e.g.
 CC neural cell differentiation). The method involves contacting cells with a
 CC neural cell regulator (i.e. a colostrinin peptide) in order to change the
 CC cells in morphology to form neural cells. Colostrinin is a proline-rich
 CC polypeptide aggregate that is present in colostrum. The method of the
 CC invention is useful for promoting the differentiation of cells and for
 CC treating damaged neural cells in a patient. The present amino acid
 CC sequence represents a specifically claimed colostrinin peptide used in
 CC the method of the invention
 XX
 SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 5; Length 18;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
 |||||
 Db 1 DQPPDVEKPDLPFQVQS 18

RESULT 6

AAM51038

ID AAM51038 standard; peptide; 18 AA.

XX

AC AAM51038;

XX

DT 30-MAY-2002 (first entry)

XX

DE Colostrinin constituent peptide.

XX

KW Colostrinin; colostrum; immunomodulator; cardiovascular;
 KW blood cell regulator; cytokine inducer; beta-casein; human.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Modified-site 18

FT /note= "optional C-terminal amidation"

XX
 PN WO200213849-A1.
 XX
 PD 21-FEB-2002.
 XX
 PF 17-AUG-2000; 2000WO-US022775.
 XX
 PR 17-AUG-2000; 2000WO-US022775.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 PA (REGE-) REGEN THERAPEUTICS PLC.
 XX
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;
 XX
 DR WPI; 2002-269150/31.
 XX
 PT Modulation of blood cell proliferation in a patient involves use of blood
 PT cell regulator selected from colostrinin, its constituent peptide and/or
 PT analog.
 XX
 PS Claim 1; Page 34; 54pp; English.
 XX
 CC The present sequence is that of a colostrinin constituent peptide that is
 CC preferred for use as an immunological regulator and as a blood cell
 CC regulator in claimed methods of the invention. It is classified as having
 CC a beta-casein homologue precursor. Methods are claimed for: inducing a
 CC cytokine in a cell by contact with an immunological regulator, where the
 CC cell is present in a cell culture, a tissue, an organ or an organism, and
 CC the cell is mammalian, including human; modulating an immune response in
 CC a cell by contact with the immunological regulator under conditions
 CC effective to induce a cytokine; modulating an immune response in a
 CC patient by administering an immunological regulator under conditions
 CC effective to induce a cytokine, where the immunological regulator is
 CC administered topically or as part of a dietary supplement, and where the
 CC immune response is specific or non specific, an interferon response or an
 CC antibody response; modulating blood cell proliferation by contacting
 CC blood cells with a blood cell regulator, where the blood cells are
 CC present in a cell culture or an organism, are mammalian or human, and
 CC where the blood cells are increased in number or differentiated; and a
 CC method for modulating blood cell proliferation in a patent. A claimed
 CC cytokine-inducing composition comprises a pharmaceutical carrier and an
 CC active agent such as the present peptide. Cytokines induced by this
 CC peptide in human leucocyte cultures include interferon-gamma, tumour
 CC necrosis factor-alpha, interleukin-6 and interleukin-10. It was one of
 CC the best overall inducers in almost all cytokine and blood cell
 CC proliferation experiments conducted
 XX
 SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 5; Length 18;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
 ||||||||||||||||
 Db 1 DQPPDVEKPDLPFQVQS 18

RESULT 7

AAE20230

ID AAE20230 standard; peptide; 18 AA.

XX

AC AAE20230;

XX

DT 18-JUN-2002 (first entry)

XX

DE Colostrinin constituent peptide #3.

XX

KW Blood cell regulator; colostrinin; constituent peptide; oxidative stress;
KW therapy; oxidative damage; skin; aging; wound healing; cell replacement;
KW tissue; organ; cosmetic procedure; repair; regeneration; preservation;
KW transplantation; implantation; dermatological; vulnerary.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Modified-site 18

FT /note= "Optionally C-terminal amide"

XX

PN WO200213850-A1.

XX

PD 21-FEB-2002.

XX

PF 17-AUG-2000; 2000WO-US022776.

XX

PR 17-AUG-2000; 2000WO-US022776.

XX

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Stanton GJ, Hughes TK, Boldogh I;

XX

DR WPI; 2002-269151/31.

XX

PT Composition useful for the modulation of blood cell proliferation in a
PT patient comprises a blood cell regulator selected from colostrinin, its
PT constituent peptide and/or analog.

XX

PS Claim 6; Page 25; 51pp; English.

XX

CC The invention relates to a composition which comprises a blood cell
CC regulator selected from colostrinin, its constituent peptide and/or
CC analogue. The invention is used for modulating the oxidative stress level
CC in a cell e.g. mammalian or human cell present in a cell culture, tissue,
CC organ, or organism; or for treating oxidative damage to the skin of a
CC patient e.g. animal or human; to modulate oxidative stress during/ after
CC a premature birth or normal birth, preventing/delaying aging in a
CC patient, enhancing wound healing, and the reduction of side effects of
CC cosmetic procedures. The method changes the level of an oxidising species
CC in the cell, such as decreases or prevents increase in the level of
CC damage to a biomolecule of the patient selected from DNA, protein and/or
CC lipid, compared to the same conditions when the oxidative stress
CC regulator is not present. The modulation of oxidative stress results in
CC enhanced repair, regeneration, and replacement of cells, tissues and
CC organs (e.g. kidney, liver, pancreas, skin, and the other internal and

CC external organs), as well as enhanced preservation of such organs for
CC transplantation, implantation, or scientific research. The present
CC sequence is a colostrinin constituent peptide
XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
| | | | | | | | | | | | | | | |
Db 1 DQPPDVEKPDLPFQVQS 18

RESULT 8
AAB59355

ID AAB59355 standard; peptide; 19 AA.

XX

AC AAB59355;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment derived sequence #15.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;
KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX

PR 02-JUN-1999; 99GB-00012852.

XX

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Georgiades JA;

XX

DR WPI; 2001-071058/08.

XX

PT Peptides having an N-terminal amino acid sequence isolated from
PT colostrinin for treating e.g. disorders of the central nervous system and
PT immune system, viral and bacterial infections, and diseases characterized
PT by amyloid plaques.

XX

PS Claim 8; Page 27; 63pp; English.

XX

CC The present invention provides the sequences of a number of peptides
CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide
CC fragment of colostrum. These peptides can be used in the treatment of
CC central nervous system disorders such as senile dementia, Parkinson's
CC disease, Alzheimer's disease, psychosis and neurosis, immune system
CC disorders such as bacterial and viral infections, to improve the

CC development of a child's immune system, as a dietary supplement, and to
CC promote the dissolution of beta-amyloid plaques
XX
SQ Sequence 19 AA;

Query Match 100.0%; Score 98; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
| | | | | | | | | | | | | | | | | |
Db 2 DQPPDVEKPDLPFQVQS 19

RESULT 9

ABU28927

ID ABU28927 standard; protein; 1047 AA.

XX

AC ABU28927;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #14454.

XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

OS Enterococcus faecalis.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

DR N-PSDB; ACA32797.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 25; SEQ ID NO 56851; 1766pp; English.

XX

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 1047 AA;

Query Match 52.0%; Score 51; DB 6; Length 1047;
 Best Local Similarity 47.1%; Pred. No. 1e+02;
 Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 DQPPDVEKPD LQPFQVQ 17
 ||| |::||: : ||::
 Db 256 DQPVDLQKPETKQFQLK 272

RESULT 10

AAM79068

ID AAM79068 standard; protein; 377 AA.

XX

AC AAM79068;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human protein SEQ ID NO 1730.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK52201.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 20; Page 4066-4067; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 377 AA;

Query Match 49.0%; Score 48; DB 4; Length 377;
 Best Local Similarity 72.7%; Pred. No. 95;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
 |||||:| ||
 Db 361 PPDVEQPQTQP 371

RESULT 11
 ABB83472
 ID ABB83472 standard; protein; 377 AA.

XX
 AC ABB83472;
 XX
 DT 30-SEP-2002 (first entry)
 XX
 DE Human cytoskeleton-associated protein, CSAP-1.
 XX
 KW Human; cytoskeleton-associated protein; CSAP; CSAP-1;
 KW cell proliferative disorder; viral infection; neurological disorder;
 KW transgenic animal; antiatherosclerotic; antipsoriatic; antiinflammatory;
 KW virucide; anticonvulsant; vasotropic; cerebroprotective; nootropic;
 KW neuroprotective; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200253719-A2.
 XX
 PD 11-JUL-2002.
 XX
 PF 04-JAN-2002; 2002WO-US000178.
 XX
 PR 04-JAN-2001; 2001US-0260085P.
 PR 13-FEB-2001; 2001US-0268554P.
 PR 14-FEB-2001; 2001US-0269111P.
 PR 23-FEB-2001; 2001US-0271211P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lu DAM, Baughn MR, Yao MG, Ding L, Honchell CD, Yue H, Tang YT;
 PI Warren BA, Duggan BM, Xu Y, Walia NK, Griffin JA, Stewart EA;
 PI Gandhi AR, Khan FA, Thangavelu K, Ison CH, Azimzai Y, Hafalia AJA;
 PI Gietzen KJ, Lal PG, Sanjanwala MM, Elliott VS;
 XX
 DR WPI; 2002-583611/62.
 DR N-PSDB; ABN85310.
 XX
 PT Novel isolated human cytoskeleton-associated protein for diagnosing,
 PT treating or preventing atherosclerosis, psoriasis, leukemia, epilepsy,
 PT ischemic cerebrovascular disease, cerebral neoplasms and Alzheimer's
 PT disease.
 XX
 PS Claim 1; Page 120-121; 167pp; English.
 XX
 CC The present sequence is the protein sequence for a human cytoskeleton-
 CC associated protein (CSAP). The CSAP and its coding sequence are useful in
 CC the diagnosis, treatment and prevention of a cell proliferative disorder
 CC such as actinic keratosis, atherosclerosis, psoriasis, primary
 CC thrombocythaemia, leukaemia; a viral infection such as those caused by
 CC adenoviruses (acute respiratory disease, pneumonia), arenaviruses
 CC (lymphocytic choriomeningitis); and a neurological disorder such as
 CC epilepsy, ischaemic cerebrovascular disease, stroke, cerebral neoplasms,
 CC Alzheimer's disease, Pick's disease, Huntington's disease or amyotrophic
 CC lateral sclerosis. The CSAP coding sequence is also useful for creating
 CC knock out or knock in humanised animals or transgenic animals to model
 CC human diseases
 XX
 SQ Sequence 377 AA;

Query Match 49.0%; Score 48; DB 5; Length 377;
Best Local Similarity 72.7%; Pred. No. 95;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
|||||:| ||
Db 361 PPDVEQPQTQP 371

RESULT 12

ADE47756

ID ADE47756 standard; protein; 377 AA.

XX

AC ADE47756;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human NOV35a protein SEQ ID NO:118.

XX

KW human; cardiant; antiarteriosclerotic; hypotensive; immunosuppressive;

KW dermatological; anorectic; cytostatic; antidiabetic; haemostatic;

KW anti-HIV; antiasthmatic; antibacterial; virucide; neuroprotective;

KW nootropic; antiparkinsonian; antilipaemic; gene therapy; vaccine.

XX

OS Homo sapiens.

XX

PN WO2003076642-A2.

XX

PD 18-SEP-2003.

XX

PF 02-AUG-2002; 2002WO-US024459.

XX

PR 02-AUG-2001; 2001US-0309501P.

PR 03-AUG-2001; 2001US-0310291P.

PR 08-AUG-2001; 2001US-0310951P.

PR 09-AUG-2001; 2001US-0311292P.

PR 13-AUG-2001; 2001US-0311979P.

PR 14-AUG-2001; 2001US-0312203P.

PR 17-AUG-2001; 2001US-0313156P.

PR 17-AUG-2001; 2001US-0313201P.

PR 20-AUG-2001; 2001US-0313702P.

PR 21-AUG-2001; 2001US-0314031P.

PR 23-AUG-2001; 2001US-0314466P.

PR 28-AUG-2001; 2001US-0315403P.

PR 29-AUG-2001; 2001US-0315853P.

PR 31-AUG-2001; 2001US-0316508P.

PR 21-SEP-2001; 2001US-0323936P.

PR 03-DEC-2001; 2001US-0338078P.

PR 05-FEB-2002; 2002US-0354655P.

PR 05-MAR-2002; 2002US-0361764P.

PR 19-APR-2002; 2002US-0373825P.

PR 15-MAY-2002; 2002US-0380971P.

PR 15-MAY-2002; 2002US-0380980P.

PR 16-MAY-2002; 2002US-0381039P.

PR 28-MAY-2002; 2002US-0383761P.

PR 29-MAY-2002; 2002US-0383887P.

PR 01-AUG-2002; 2002US-00210130.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Zerhusen BD, Patturajan M, Kekuda R, Miller CE, Rieger DK;
PI Pena CEA, Shimkets RA, Li L, Berghs C, Zhong M, Casman SJ, Voss EZ;
PI Boldog FL, Padigar M, Smithson G, Shenoy SG, Ji W, Gorman L;
PI Vernet CAM, Leite MW, Guo X, Anderson DW, Spytek KA, Gerlach VL;
PI Burgess CE, Khramtsov NV, Ort T, Ellerman K, Rastelli L, Agee ML;
PI Chaudhuri A, Chant JS, Dipippo VA, Edinger SR, Eisen A, Gangolli EA;
PI Giot L, Ooi CE, Rothenberg ME, Spaderna SK, Hjalt T, Liu X;
PI Taupier RJ, Catterton E;

XX

DR WPI; 2003-779062/73.

DR N-PSDB; ADE47755.

XX

PT New NOVX polypeptides and nucleic acids, useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, diabetes,
PT atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing
PT or pharmacogenomics.

XX

PS Claim 1; SEQ ID NO 118; 562pp; English.

XX

CC The invention relates to a novel (NOVX) human polypeptide. A polypeptide
CC of the invention has cardiant, antiarteriosclerotic, hypotensive,
CC immunosuppressive, dermatological, anorectic, cytostatic, antidiabetic,
CC haemostatic, anti-HIV, antiasthmatic, antibacterial, virucide,
CC neuroprotective, nootropic, antiparkinsonian, and antilipaemic activity.
CC A polynucleotide encoding a polypeptide of the invention may have a use
CC in gene therapy, and as a vaccine. A polypeptide of the invention is
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, the disease selected from a pathology
CC associated with the polypeptide. These may also be used in diagnosing,
CC treating or preventing NOVX-associated disorders such as cardiomyopathy,
CC atherosclerosis, hypertension, scleroderma, obesity, cancer, diabetes,
CC haemophilia, graft-versus-host disease, AIDS, asthma, Crohn's disease,
CC multiple sclerosis, infections, anorexia, cancer-associated cachexia,
CC neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's
CC disease), haematopoietic disorders, dyslipidaemias and other wasting
CC disorders associated with chronic diseases. The nucleic acids are also
CC used as hybridisation probes, in chromosome mapping, tissue typing,
CC preventive medicine, and pharmacogenomics. The polypeptides are also
CC useful as vaccines. The present sequence represents a NOVX polypeptide of
CC the invention.

XX

SQ Sequence 377 AA;

Query Match 49.0%; Score 48; DB 7; Length 377;

Best Local Similarity 72.7%; Pred. No. 95;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPDVEKPDLP 13

||||:| ||

Db 361 PPDVEQPQTQP 371

RESULT 13

AAU35153

ID AAU35153 standard; protein; 541 AA.

XX

AC AAU35153;

XX

DT 13-FEB-2002 (first entry)

XX

DE Enterococcus faecalis cellular proliferation protein #440.

XX

KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.

XX

OS Enterococcus faecalis.

XX

PN WO200170955-A2.

XX

PD 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-US009180.

XX

PR 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;

XX

DR WPI; 2001-611495/70.

DR N-PSDB; AAS53012.

XX

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.

XX

PS Example 3; SEQ ID NO 10746; 511pp; English.

XX

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 541 AA;

Query Match 49.0%; Score 48; DB 4; Length 541;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQV 16
|| | || ||| :|
Db 274 DQQPGKEKWDLPMEV 289

RESULT 14

ABB91350

ID ABB91350 standard; protein; 516 AA.

XX

AC

ABB91350;

XX

DT

31-MAY-2002 (first entry)

XX

DE

Herbicidally active polypeptide SEQ ID NO 561.

XX

KW

Herbicide; plant; agriculture; herbicide.

XX

OS

Arabidopsis thaliana.

XX

PN

WO200210210-A2.

XX

PD

07-FEB-2002.

XX

PF

28-AUG-2001; 2001WO-EP009892.

XX

PR

28-AUG-2001; 2001WO-EP009892.

XX

PA

(FARB) BAYER AG.

XX

PI

Tietjen K, Weidler M;

XX

DR

WPI; 2002-269010/31.

XX

PT

Identifying plant target proteins for herbicidally active compounds,
comprising aligning and comparing nucleic acid or amino acid sequences
from plant with nucleic acid or amino acid sequences from non-plant
organisms.

XX

PS

Claim 5; SEQ ID NO 561; 261pp + Sequence Listing; English.

XX

CC

The invention relates to identifying target proteins (ABB90790-ABB94016)
for herbicidally active compounds, comprising aligning and comparing
nucleic acid or amino acid sequences from plant with nucleic acid or
amino acid sequences from non-plant organisms using suitable search
parameters, where plant sequences having an E-value greater by a factor
of 3 than the E-value of most similar non-plant sequences are selected.
The polypeptides or nucleic acids encoding them are useful for
identifying modulators. The identified modulators are useful as

CC herbicides

XX

SQ Sequence 516 AA;

Query Match 48.0%; Score 47; DB 5; Length 516;

Best Local Similarity 56.2%; Pred. No. 1.8e+02;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPQPFQVQS 18

| | : | | | | | : |

Db 310 PMDIEKTDNQPFETLAS 325

RESULT 15

ABB91351

ID ABB91351 standard; protein; 719 AA.

XX

AC ABB91351;

XX

DT 31-MAY-2002 (first entry)

XX

DE Herbicidally active polypeptide SEQ ID NO 562.

XX

KW Herbicidal; plant; agriculture; herbicide.

XX

OS Arabidopsis thaliana.

XX

PN WO200210210-A2.

XX

PD 07-FEB-2002.

XX

PF 28-AUG-2001; 2001WO-EP009892.

XX

PR 28-AUG-2001; 2001WO-EP009892.

XX

PA (FARB) BAYER AG.

XX

PI Tietjen K, Weidler M;

XX

DR WPI; 2002-269010/31.

XX

PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.

XX

PS Claim 5; SEQ ID NO 562; 261pp + Sequence Listing; English.

XX

CC The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides

XX

SQ Sequence 719 AA;

Query Match 48.0%; Score 47; DB 5; Length 719;

Best Local Similarity 56.2%; Pred. No. 2.6e+02;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQVQS 18

| | : | | | | : |

Db 324 PMDIEKTDNQPFLLAS 339

Search completed: August 24, 2004, 15:42:20

Job time : 82.3433 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:33:13 ; Search time 19.7463 Seconds
(without alignments)
47.060 Million cell updates/sec

Title: US-09-641-801-3
Perfect score: 98
Sequence: 1 DQPPDVEKPDLPFQVQS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	98	100.0	18	4	US-09-641-803-3	Sequence 3, Appli
2	51	52.0	1056	4	US-09-134-000C-5086	Sequence 5086, Ap
3	48	49.0	400	4	US-09-134-000C-4389	Sequence 4389, Ap
4	46	46.9	180	4	US-09-744-128-17	Sequence 17, Appl
5	46	46.9	194	4	US-09-744-128-16	Sequence 16, Appl
6	45	45.9	26	4	US-09-406-781-13	Sequence 13, Appl
7	45	45.9	26	4	US-09-880-132-13	Sequence 13, Appl
8	43	43.9	1462	3	US-07-792-600-31	Sequence 31, Appl
9	43	43.9	1462	3	US-09-157-021-31	Sequence 31, Appl
10	43	43.9	1462	3	US-09-156-842-31	Sequence 31, Appl
11	43	43.9	1462	4	US-09-591-514-31	Sequence 31, Appl

12	42	42.9	460	1	US-08-630-592-7	Sequence 7, Appli
13	42	42.9	460	1	US-08-714-991-7	Sequence 7, Appli
14	42	42.9	460	3	US-09-032-365A-8	Sequence 8, Appli
15	42	42.9	506	1	US-08-631-200-8	Sequence 8, Appli
16	42	42.9	506	2	US-08-829-553-8	Sequence 8, Appli
17	42	42.9	506	2	US-08-922-267A-8	Sequence 8, Appli
18	42	42.9	506	2	US-08-936-707A-8	Sequence 8, Appli
19	42	42.9	506	2	US-08-936-706A-8	Sequence 8, Appli
20	42	42.9	506	3	US-09-248-203-8	Sequence 8, Appli
21	42	42.9	506	3	US-09-032-365A-62	Sequence 62, Appl
22	42	42.9	506	3	US-08-812-824-4	Sequence 4, Appli
23	42	42.9	506	3	US-09-406-071-8	Sequence 8, Appli
24	42	42.9	506	3	US-08-955-918C-10	Sequence 10, Appl
25	42	42.9	506	4	US-08-697-766A-10	Sequence 10, Appl
26	42	42.9	506	4	US-09-814-986-8	Sequence 8, Appli
27	42	42.9	512	3	US-09-032-365A-60	Sequence 60, Appl
28	42	42.9	518	3	US-09-032-365A-58	Sequence 58, Appl
29	42	42.9	561	1	US-08-714-991-27	Sequence 27, Appl
30	42	42.9	561	3	US-09-032-365A-10	Sequence 10, Appl
31	42	42.9	604	4	US-09-489-039A-8519	Sequence 8519, Ap
32	41.5	42.3	553	4	US-09-252-991A-22589	Sequence 22589, A
33	41	41.8	27	2	US-08-563-892A-2	Sequence 2, Appli
34	41	41.8	27	2	US-08-563-892A-23	Sequence 23, Appl
35	41	41.8	27	2	US-08-563-892A-24	Sequence 24, Appl
36	41	41.8	27	2	US-08-563-892A-25	Sequence 25, Appl
37	41	41.8	27	4	US-09-366-212A-2	Sequence 2, Appli
38	41	41.8	27	4	US-09-366-212A-23	Sequence 23, Appl
39	41	41.8	27	4	US-09-366-212A-24	Sequence 24, Appl
40	41	41.8	27	4	US-09-366-212A-25	Sequence 25, Appl
41	41	41.8	642	4	US-08-461-562B-9	Sequence 9, Appli
42	41	41.8	2332	1	US-07-864-004B-4	Sequence 4, Appli
43	41	41.8	2332	1	US-08-251-937A-4	Sequence 4, Appli
44	41	41.8	2332	1	US-08-212-133A-2	Sequence 2, Appli
45	41	41.8	2332	1	US-08-276-594A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-641-803-3

; Sequence 3, Application US/09641803

; Patent No. 6500798

; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John

; APPLICANT: HUGHES, Thomas K.

; APPLICANT: BOLDOGH, Istvan

; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND

; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS

; FILE REFERENCE: 265.00220101

; CURRENT APPLICATION NUMBER: US/09/641,803

; CURRENT FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: 60/149,310

; PRIOR FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: peptide
US-09-641-803-3

Query Match 100.0%; Score 98; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 9.8e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFFQVQS 18
| | | | | | | | | | | | | | | |
Db 1 DQPPDVEKPDLPFFQVQS 18

RESULT 2

US-09-134-000C-5086
; Sequence 5086, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5086
; LENGTH: 1056
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5086

Query Match 52.0%; Score 51; DB 4; Length 1056;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFFQVQ 17
| | | | : : : : : : : : : :
Db 265 DQPVDLQKPETKQFQLK 281

RESULT 3

US-09-134-000C-4389
; Sequence 4389, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4389
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4389

Query Match 49.0%; Score 48; DB 4; Length 400;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQV 16
||| ||||| :|
Db 133 DQQPGKEKWDLPMEV 148

RESULT 4

US-09-744-128-17
; Sequence 17, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Veis et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 17
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequence: PCR product
US-09-744-128-17

Query Match 46.9%; Score 46; DB 4; Length 180;
Best Local Similarity 50.0%; Pred. No. 9.4;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQVQS 18
|| :|| ||||| :|
Db 105 PPSAQQPFQPFQQA 120

RESULT 5

US-09-744-128-16
; Sequence 16, Application US/09744128
; Patent No. 6677306


```
; GENERAL INFORMATION:
; APPLICANT: Veis et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 16
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequence: PCR product
US-09-744-128-16
```

```
Query Match          46.9%; Score 46; DB 4; Length 194;
Best Local Similarity 50.0%; Pred. No. 10;
Matches      8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy          3 PPDVEKPDLPFQVQS 18
             ||  ::|  |||  |:
Db          119 PPSAQQPFQPFQQA 134
```

```
RESULT 6
US-09-406-781-13
; Sequence 13, Application US/09406781
; Patent No. 6306663
; GENERAL INFORMATION:
; APPLICANT: Kenten, John
; APPLICANT: Roberts, Steven
; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUARYOTIC ORGANISMS
; FILE REFERENCE: 2757-3
; CURRENT APPLICATION NUMBER: US/09/406,781
; CURRENT FILING DATE: 1999-09-28
; EARLIER APPLICATION NUMBER: 60/119,851
; EARLIER FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: PEST example
; OTHER INFORMATION: sequence
US-09-406-781-13
```

```
Query Match          45.9%; Score 45; DB 4; Length 26;
Best Local Similarity 57.1%; Pred. No. 1.6;
Matches      8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

Qy 3 PPDVEKPDLPFQV 16
|| ||: ||: | |
Db 2 PPGVEEPDVGPLPV 15

RESULT 7

US-09-880-132-13
; Sequence 13, Application US/09880132
; Patent No. 6559280
; GENERAL INFORMATION:
; APPLICANT: Kenten, John
; APPLICANT: Roberts, Steven
; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUCARYOTIC ORGANISMS
; FILE REFERENCE: 2757-6
; CURRENT APPLICATION NUMBER: US/09/880,132
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 09/406,781
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: 60/119,851
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: PEST example
; OTHER INFORMATION: sequence
US-09-880-132-13

Query Match 45.9%; Score 45; DB 4; Length 26;
Best Local Similarity 57.1%; Pred. No. 1.6;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQV 16
|| ||: ||: | |
Db 2 PPGVEEPDVGPLPV 15

RESULT 8

US-07-792-600-31
; Sequence 31, Application US/07792600
; Patent No. 6008045
; GENERAL INFORMATION:
; APPLICANT: COPELAND, WILLIAM C.
; APPLICANT: WANG, TERESA S.-F.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: TEMPLATE-DEPENDENT ENZYMATIc SYNTHESIS OF NUCLEIC ACID
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter G. Carroll
; STREET: 220 Montgomery Street, Suite 710
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94104

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,600
; FILING DATE: 19911115
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: STDU-00097
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1462 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-792-600-31

```

```

Query Match          43.9%; Score 43; DB 3; Length 1462;
Best Local Similarity 38.9%; Pred. No. 2.7e+02;
Matches      7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      1 DQPPDVEKPDLPFQVQS 18
        |:| :||: ||:|  ::
Db      255 DEPMEVEEVDLEPMAAKA 272

```

```

RESULT 9
US-09-157-021-31
; Sequence 31, Application US/09157021A
; Patent No. 6100023
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/157,021A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1462
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-157-021-31

```

```

Query Match          43.9%; Score 43; DB 3; Length 1462;
Best Local Similarity 38.9%; Pred. No. 2.7e+02;
Matches      7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

```

Qy 1 DQPPDVEKPDLPFQVQS 18
|:| :||: ||:| ::
Db 255 DEPMEVEEVDLEPMAAKA 272

RESULT 10

US-09-156-842-31
; Sequence 31, Application US/09156842A
; Patent No. 6103473
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Screening
; FILE REFERENCE: STDU-03485
; CURRENT APPLICATION NUMBER: US/09/156,842A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1462
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-156-842-31

Query Match 43.9%; Score 43; DB 3; Length 1462;
Best Local Similarity 38.9%; Pred. No. 2.7e+02;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQS 18
|:| :||: ||:| ::
Db 255 DEPMEVEEVDLEPMAAKA 272

RESULT 11

US-09-591-514-31
; Sequence 31, Application US/09591514
; Patent No. 6670161
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/591,514
; CURRENT FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US/09/157,021
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 07/792,600
; PRIOR FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1462
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-591-514-31

Query Match 43.9%; Score 43; DB 4; Length 1462;
Best Local Similarity 38.9%; Pred. No. 2.7e+02;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFFQVQS 18
|:| :||: ||:| ::
Db 255 DEPMEVEEVDLEPMAAKA 272

RESULT 12

US-08-630-592-7

; Sequence 7, Application US/08630592
; Patent No. 5770432
; GENERAL INFORMATION:
; APPLICANT: Nishina, Patsy
; APPLICANT: No. 5770432enTrauth, Konrad
; APPLICANT: Naggert, Juerger
; APPLICANT: No. 5770432th, Michael
; TITLE OF INVENTION: Obesity Associated Genes
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 941114187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PCDOS/MSDOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,592
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A59504/BIR/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 7811989
; TELEFAX: (415) 3983249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-592-7

Query Match 42.9%; Score 42; DB 1; Length 460;
Best Local Similarity 47.1%; Pred. No. 1.1e+02;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
:|| ||| ||: | ::
Db 195 EQPVDVEVQDLEEFALR 211

RESULT 13

US-08-714-991-7

; Sequence 7, Application US/08714991
; Patent No. 5776762
; GENERAL INFORMATION:
; APPLICANT: NORTH, Michael
; APPLICANT: NISHINA, Patsy
; APPLICANT: No. 5776762en-Trauth, Konrad
; APPLICANT: NAGGERT, Juergen
; TITLE OF INVENTION: OBESITY ASSOCIATED GENES
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,991
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A-59504-1/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-8700
; TELEFAX: 415-494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-714-991-7

Query Match 42.9%; Score 42; DB 1; Length 460;
Best Local Similarity 47.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
:|| ||| ||: | ::

Db 195 EQPVDVEVQDLEEFALR 211

RESULT 14

US-09-032-365A-8

; Sequence 8, Application US/09032365A
; Patent No. 6114502
; GENERAL INFORMATION:
; APPLICANT: No. 6114502th, Michael
; APPLICANT: Nishina, Patsy
; APPLICANT: Naggart, Juergen
; APPLICANT: No. 6114502en-Trauth, Konrad
; TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
; TITLE OF INVENTION: NEUROSENSORY DEFECTS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,365A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-2CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650 327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-032-365A-8

Query Match 42.9%; Score 42; DB 3; Length 460;
Best Local Similarity 47.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
: || ||| ||: | ::

Db 195 EQPVDVEVQDLEEFALR 211

RESULT 15

US-08-631-200-8

; Sequence 8, Application US/08631200
; Patent No. 5646040
; GENERAL INFORMATION:
; APPLICANT: Kleyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/631,200
; FILING DATE: 12-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 506 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

US-08-631-200-8

Query Match 42.9%; Score 42; DB 1; Length 506;
Best Local Similarity 47.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
: || ||| ||: | ::
Db 241 EQPVDVEVQDLEEFALR 257

Search completed: August 24, 2004, 15:55:13
Job time : 21.7463 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:26:28 ; Search time 17.4627 Seconds
(without alignments)
99.151 Million cell updates/sec

Title: US-09-641-801-3
Perfect score: 98
Sequence: 1 DQPPDVEKPDLPFQVQS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	55	56.1	483	2	D87543	methylnalonyl-CoA
2	47	48.0	404	2	D42094	bw4 protein - smut
3	47	48.0	719	2	F96577	hypothetical prote
4	46	46.9	134	2	T36365	proline-rich prote
5	46	46.9	154	2	I49486	amelogenin (Enamel
6	46	46.9	187	2	AH2802	conserved hypothet
7	46	46.9	196	2	PC1148	amelogenin precurs
8	46	46.9	196	2	JC2391	amelogenin precurs
9	46	46.9	199	2	A97582	hypothetical prote
10	46	46.9	728	2	A75582	serine proteinase,
11	45	45.9	858	2	A57513	heat shock protein
12	45	45.9	859	2	S66666	heat shock protein
13	45	45.9	5170	2	T15348	hypothetical prote

14	44	44.9	175	2	E85078	hypothetical prote
15	44	44.9	380	2	T48491	gibberellin 20-oxi
16	44	44.9	1070	2	JC4852	DNA ligase (ATP) (
17	44	44.9	1182	2	T30189	myelin transcripti
18	44	44.9	1188	2	T46608	zinc finger protei
19	43	43.9	211	2	AD1053	conserved hypothet
20	43	43.9	335	2	T34837	probable transfera
21	43	43.9	699	2	I38073	nucleolar phosphop
22	43	43.9	1462	1	DJHUAC	DNA-directed DNA p
23	42.5	43.4	346	2	S58326	hypothetical prote
24	42.5	43.4	595	2	T41042	hypothetical prote
25	42.5	43.4	2774	2	A43359	microtubule-associ
26	42	42.9	228	2	S46965	microfilarial shea
27	42	42.9	340	2	AE3562	fructose-bisphosph
28	42	42.9	358	1	S25464	transcription fact
29	42	42.9	372	2	A46037	Hox-1.11 - mouse
30	42	42.9	393	2	T49578	hypothetical prote
31	42	42.9	396	2	T27946	hypothetical prote
32	42	42.9	460	2	JC2194	vasoactive intesti
33	42	42.9	495	2	JC2195	vasoactive intesti
34	42	42.9	499	2	T42930	hypothetical prote
35	42	42.9	547	2	H87679	sensor histidine k
36	42	42.9	640	2	S23008	insulin-like growt
37	42	42.9	1047	2	T25782	hypothetical prote
38	42	42.9	1138	2	T24635	hypothetical prote
39	42	42.9	1199	2	T18348	probable pol polyp
40	42	42.9	2500	2	G71609	hypothetical prote
41	41.5	42.3	710	2	F86290	hypothetical prote
42	41.5	42.3	717	2	S50852	cleavage stimulati
43	41	41.8	70	2	A48841	dynamamin - bovine (
44	41	41.8	148	1	QQBE20	BLLF2 protein - hu
45	41	41.8	189	2	A31348	T-cell receptor CD

ALIGNMENTS

RESULT 1

D87543

methyalmalonyl-CoA mutase, beta subunit [imported] - *Caulobacter crescentus*

C;Species: *Caulobacter crescentus*

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 23-Sep-2002

C;Accession: D87543

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: D87543

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-483 <STO>

A;Cross-references: GB:AE005673; NID:g13423904; PIDN:AAK24344.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2373
C;Superfamily: methylmalonyl-CoA mutase beta chain

Query Match 56.1%; Score 55; DB 2; Length 483;
Best Local Similarity 58.8%; Pred. No. 0.85;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
|:|:| | | | |
Db 439 DKPPEVETPDSSAFVQ 455

RESULT 2

D42094

bw4 protein - smut fungus (*Ustilago maydis*) (fragment)

C;Species: *Ustilago maydis* (corn smut)

C;Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 24-Sep-1999

C;Accession: D42094

R;Gillissen, B.; Bergemann, J.; Sandmann, C.; Schroeer, B.; Boelker, M.;
Kahmann, R.

Cell 68, 647-657, 1992

A;Title: A two-component regulatory system for self/non-self recognition in
Ustilago maydis.

A;Reference number: A42094; MUID:92154679; PMID:1739973

A;Accession: D42094

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-404 <GIL>

A;Cross-references: GB:M84181; NID:g170578; PIDN:AAA34223.1; PID:g170579

C;Superfamily: unassigned homeobox proteins; homeobox homology

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

F;136-192/Domain: homeobox homology <HOX>

Query Match 48.0%; Score 47; DB 2; Length 404;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 QPPDVEKPDLPFQVQS 18
:| | :| | | | :|
Db 198 EPTDSTQPDLSFRSES 214

RESULT 3

F96577

hypothetical protein F22G10.3 [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 24-Aug-2001

C;Accession: F96577

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;

Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;

Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;

Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;

Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;

Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzialli, A.; Militischer, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F96577

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-719 <STO>

A;Cross-references: GB:AE005173; NID:g10645349; PIDN:AAG21469.1; GSPDB:GN00141
C;Genetics:

A;Gene: F22G10.3

A;Map position: 1

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology

Query Match	48.0%;	Score 47;	DB 2;	Length 719;
Best Local Similarity	56.2%;	Pred. No. 24;		
Matches	9;	Conservative	2;	Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQVQS 18
| | : | | | : |
Db 324 PMDIEKTDNQPFTLAS 339

RESULT 4

T36365

proline-rich protein SCE94.05 - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T36365

R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A;Reference number: Z21573

A;Accession: T36365

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-134 <OLI>

A;Cross-references: EMBL:AL049628; PIDN:CAB40854.1; GSPDB:GN00070;
SCOEDB:SCE94.05

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SCE94.05

Query Match	46.9%;	Score 46;	DB 2;	Length 134;
Best Local Similarity	57.1%;	Pred. No. 5;		
Matches	8;	Conservative	1;	Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPF 14

| ||| || :||
Db 34 DPPDSPFPDPEPF 47

RESULT 5

I49486

amelogenin (Enamel-Specific Protein) - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Aug-1999

C;Accession: I49486

R;Snead, M.L.; Lau, E.C.; Zeichner-David, M.; Fincham, A.G.; Woo, S.L.C.; Slavkin, H.C.

Biochem. Biophys. Res. Commun. 129, 812-818, 1985

A;Title: DNA sequence for cloned cDNA for murine amelogenin reveal the amino acid sequence for enamel-specific protein.

A;Reference number: I49486; MUID:85251692; PMID:4015654

A;Accession: I49486

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-154 <RES>

A;Cross-references: GB:M10095; NID:g191894; PIDN:AAA37218.1; PID:g191895

C;Superfamily: amelogenin

Query Match 46.9%; Score 46; DB 2; Length 154;
Best Local Similarity 50.0%; Pred. No. 5.9;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQVQS 18

 || :|| ||| |:

Db 79 PPSAQPFQPFQQA 94

RESULT 6

AH2802

conserved hypothetical protein Atu1842 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C;Accession: AH2802

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee Sr., D.; Chapman, P.; Clendenning, J.; Deatherage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyaivin, T.; Levy, R.; Li, M.; McClelland, E.; Palmieri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AH2802

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-187 <KUR>

A;Cross-references: GB:AE008688; PIDN:AAL42838.1; PID:g17740287; GSPDB:GN00186

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: Atul842

A;Map position: circular chromosome

Query Match 46.9%; Score 46; DB 2; Length 187;
Best Local Similarity 63.6%; Pred. No. 7.4;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
|||| ||::|
Db 95 PPDVANPDIRP 105

RESULT 7

PC1148

amelogenin precursor - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

C;Accession: PC1148

R;Lau, E.C.; Simmer, J.P.; Bringas Jr., P.; Hsu, D.D.J.; Hu, C.C.; Zeichner-David, M.; Thiemann, F.; Snead, M.L.; Slavkin, H.C.; Fincham, A.G.
Biochem. Biophys. Res. Commun. 188, 1253-1260, 1992

A;Title: Alternative splicing of the mouse amelogenin primary RNA transcript contributes to amelogenin heterogeneity.

A;Reference number: PC1148; MUID:93075222; PMID:1445358

A;Accession: PC1148

A;Molecule type: mRNA

A;Residues: 1-196 <LAU>

C;Superfamily: amelogenin

C;Keywords: enamel; phosphoprotein; tooth

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-196/Product: amelogenin (fragment) #status predicted <MAT>

Query Match 46.9%; Score 46; DB 2; Length 196;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQVQS 18
|| ::| ||| |:
Db 121 PPSAQQPFQPFQQA 136

RESULT 8

JC2391

amelogenin precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999

C;Accession: JC2391; A45914; S50218

R;Bonass, W.A.; Robinson, P.A.; Kirkham, J.; Shore, R.C.; Robinson, C.
Biochem. Biophys. Res. Commun. 198, 755-763, 1994

A;Title: Molecular cloning and DNA sequence of rat amelogenin and a comparative analysis of mammalian amelogenin protein sequence divergence.

A;Reference number: JC2391; MUID:94128126; PMID:8297387

A;Accession: JC2391

A;Molecule type: mRNA

A;Residues: 1-196 <BON>

A;Cross-references: EMBL:U01245; NID:g415627; PIDN:AAA20491.1; PID:g521104
R;Hubbard, M.J.
submitted to the Protein Sequence Database, April 1993
A;Reference number: A45914
A;Accession: A45914
A;Status: preliminary
A;Molecule type: protein
A;Residues: 17-99,'X',101-102,'X',104-106,'X',108-109,'XXT' <HUB>
R;Bonass, W.A.; Kirkham, J.; Brookes, S.J.; Shore, R.C.; Robinson, C.
Biochim. Biophys. Acta 1219, 690-692, 1994
A;Title: Isolation and characterisation of an alternatively-spliced rat
amelogenin cDNA: LRAP - a highly conserved, functional alternatively-spliced
amelogenin?
A;Reference number: S50218; MUID:95035099; PMID:7948026
A;Accession: S50218
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-49,171-196 <BO2>
A;Cross-references: EMBL:U07054; NID:g460925; PIDN:AAA61964.1; PID:g521108
C;Genetics:
A;Introns: 18/3; 34/3; 49/3; 195/3
C;Superfamily: amelogenin
C;Keywords: alternative splicing
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-196/Product: amelogenin #status predicted <MAT>

Query Match 46.9%; Score 46; DB 2; Length 196;
Best Local Similarity 50.0%; Pred. No. 7.8;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQVQS 18
|| ::| ||| |:
Db 121 PPSAQQPFQPFQQA 136

RESULT 9

A97582

hypothetical protein AGR_C_3379 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)

C;Species: Agrobacterium tumefaciens

C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C;Accession: A97582

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, B.S.; Cao, Y.; Askenazi, M.; Halling, C.; Mullin, L.; Houmiel, K.; Gordon, J.; Vaudin, M.; Iartchouk, O.; Epp, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Flanagan, C.; Crowell, C.; Gurson, J.; Lomo, C.; Sear, C.; Strub, G.; Cielo, C.; Slater, S.
Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58.

A;Reference number: A97359; MUID:21608551; PMID:11743194

A;Accession: A97582

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-199 <KUR>

A;Cross-references: GB:AE007869; PIDN:AAK87610.1; PID:g15156956; GSPDB:GN00169

C;Genetics:

A;Gene: AGR_C_3379
A;Map position: circular chromosome

Query Match 46.9%; Score 46; DB 2; Length 199;
Best Local Similarity 63.6%; Pred. No. 7.9;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
|||| ||::|
Db 107 PPDVANPDIRP 117

RESULT 10

A75582

serine proteinase, subtilase family - *Deinococcus radiodurans* (strain R1)

C;Species: *Deinococcus radiodurans*

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C;Accession: A75582

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.;
Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; Moffat,
K.S.; Qin, H.; Jiang, L.; Pamphile, W.; Crosby, M.; Shen, M.; Vamathevan, J.J.;
Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Makarova, K.S.; Aravind, L.;
Daly, M.J.; Minton, K.W.; Fleischmann, R.D.; Ketchum, K.A.; Nelson, K.E.;
Salzberg, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans*
R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: A75582

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-728 <WHI>

A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12479.1;

PID:g6460774; TIGR:DRA0283; GSPDB:GN00078

A;Experimental source: strain R1

C;Genetics:

A;Gene: DRA0283

A;Map position: 2

Query Match 46.9%; Score 46; DB 2; Length 728;
Best Local Similarity 61.5%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLP 13
| | | | |||:
Db 188 DDPSPDVSHPDLP 200

RESULT 11

A57513

heat shock protein 110k - Chinese hamster

C;Species: *Cricetulus griseus* (Chinese hamster)

C;Date: 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change 29-Sep-1999

C;Accession: A57513; S51311

R;Lee-Yoon, D.; Easton, D.; Murawski, M.; Burd, R.; Subjeck, J.R.

J. Biol. Chem. 270, 15725-15733, 1995

A;Title: Identification of a major subfamily of large hsp70-like proteins through the cloning of the mammalian 110-kDa heat shock protein.
 A;Reference number: A57513; MUID:95318163; PMID:7797574
 A;Accession: A57513
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-858 <LEE>
 A;Cross-references: GB:Z47807; NID:g633180; PIDN:CAA87768.1; PID:g633181
 R;Yoon, D.; Murawski, M.J.; Burd, R.; Easton, D.P.; Subjeck, J.R.
 submitted to the EMBL Data Library, January 1995
 A;Description: Identification of a major subfamily of large hsp70-like proteins through the cloning of the mammalian 110 kDa heat shock protein.
 A;Reference number: S51311
 A;Accession: S51311
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-858 <YOO>
 A;Cross-references: EMBL:Z47807; NID:g633180; PIDN:CAA87768.1; PID:g633181
 C;Superfamily: heat shock protein 91
 C;Keywords: heat shock; stress-induced protein

Query Match	45.9%;	Score 45;	DB 2;	Length 858;
Best Local Similarity	41.2%;	Pred. No. 60;		
Matches	7;	Conservative	5;	Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
 | | | | : : | | : : | :
 Db 580 DQPPEAKKPKIKVWVE 596

RESULT 12
 S66666
 heat shock protein (clone E7I) - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 29-Sep-1999
 C;Accession: S66666; S72507
 R;Morozov, A.; Subjeck, J.; Raychaudhuri, P.
 FEBS Lett. 371, 214-218, 1995
 A;Title: HPV16 E7 oncoprotein induces expression of a 110 kDa heat shock protein.
 A;Reference number: S66666; MUID:96013135; PMID:7556594
 A;Accession: S66666
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-859 <MOR>
 A;Cross-references: EMBL:L40406
 R;Morozov, A.; Subjeck, J.; Raychaudhuri, P.
 submitted to the EMBL Data Library, June 1995
 A;Reference number: S72507
 A;Accession: S72507
 A;Molecule type: mRNA
 A;Residues: 1-182,'I',184-450,'G',452-493,'T',495-575,'NE',578-615,'Y',617-629,'E',631-757,'N',759-859 <MOW>
 A;Cross-references: EMBL:L40406; NID:g840651; PIDN:AAA99485.1; PID:g840652
 C;Genetics:
 A;Gene: hsp-E7I
 C;Superfamily: heat shock protein 91

C;Keywords: heat shock; molecular chaperone; stress-induced protein

Query Match 45.9%; Score 45; DB 2; Length 859;
Best Local Similarity 41.2%; Pred. No. 60;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
||||: ||| :: |:
Db 581 DQPPEAKKPKIKVVNVE 597

RESULT 13

T15348

hypothetical protein B0350.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C;Accession: T15348

R;Gattung, S.

submitted to the EMBL Data Library, February 1996

A;Description: The sequence of *C. elegans* cosmid B0350.

A;Reference number: Z18332

A;Accession: T15348

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-5170 <GAT>

A;Cross-references: EMBL:U50071; NID:g1208871; PID:g1208877; PIDN:AAA93447.1;

CESP:B0350.1

C;Genetics:

A;Gene: CESP:B0350.1

A;Introns: 48/1; 5039/3; 5116/3

Query Match 45.9%; Score 45; DB 2; Length 5170;
Best Local Similarity 66.7%; Pred. No. 4.6e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPQ 12
: || | |||||:
Db 2172 EQPHDEEKPDLE 2183

RESULT 14

E85078

hypothetical protein AT4g07990 [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C;Accession: E85078

R;anonymous, The European Union *Arabidopsis* Genome Sequencing Consortium, The Cold Spring Harbor, Washington University in St Louis and PE Biosystems *Arabidopsis* Sequencing Consortium.

Nature 402, 769-777, 1999

A;Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A;Reference number: A85001; MUID:20083488; PMID:10617198

A;Accession: E85078

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-175 <STO>

A;Cross-references: GB:NC_001268; NID:g7267438; PIDN:CAB77950.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g07990
A;Map position: 4

Query Match 44.9%; Score 44; DB 2; Length 175;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQV 16
|| :||| :|::|
Db 131 PPPQQKPDSPWEV 144

RESULT 15

T48491
gibberellin 20-oxidase - Arabidopsis thaliana
N;Alternate names: protein T28J14.140
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C;Accession: T48491
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.;
Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T48491
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-380 <BEV>
A;Cross-references: EMBL:AL163652
A;Experimental source: cultivar Columbia; BAC clone T28J14
C;Genetics:
A;Map position: 5
A;Introns: 183/2; 290/3
A;Note: T28J14.140
C;Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 44.9%; Score 44; DB 2; Length 380;
Best Local Similarity 66.7%; Pred. No. 34;
Matches 10; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

Qy 4 PDVEKP--DLQPFQV 16
|| ||| |:|| ||
Db 44 PDHEKPSTDVQPLQV 58

Search completed: August 24, 2004, 15:52:44
Job time : 20.4627 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:51:19 ; Search time 65.1493 Seconds
(without alignments)
86.825 Million cell updates/sec

Title: US-09-641-801-3
Perfect score: 98
Sequence: 1 DQPPDVEKPDLQPFQVQS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query				Description
No.	Score	Match Length	DB	ID		

1	98	100.0	18	14	US-10-281-652-3	Sequence 3, Appli
2	51	52.0	1047	12	US-10-282-122A-56851	Sequence 56851, A
3	48	49.0	377	15	US-10-210-130-118	Sequence 118, App
4	48	49.0	377	16	US-10-250-613-1	Sequence 1, Appli
5	48	49.0	541	9	US-09-815-242-10746	Sequence 10746, A
6	47	48.0	162	16	US-10-437-963-204178	Sequence 204178,
7	46	46.9	212	16	US-10-437-963-200861	Sequence 200861,
8	46	46.9	313	12	US-10-425-114-52231	Sequence 52231, A
9	46	46.9	336	12	US-10-424-599-233789	Sequence 233789,
10	46	46.9	337	12	US-10-425-114-39945	Sequence 39945, A
11	45	45.9	26	9	US-09-880-149-13	Sequence 13, Appl
12	45	45.9	26	9	US-09-880-132-13	Sequence 13, Appl
13	45	45.9	26	14	US-10-345-281-13	Sequence 13, Appl
14	45	45.9	159	12	US-10-424-599-228647	Sequence 228647,
15	45	45.9	848	9	US-09-925-302-766	Sequence 766, App
16	45	45.9	848	12	US-09-925-302-766	Sequence 766, App
17	45	45.9	1006	16	US-10-437-963-136358	Sequence 136358,
18	45	45.9	1289	16	US-10-437-963-157119	Sequence 157119,
19	44.5	45.4	583	16	US-10-437-963-114326	Sequence 114326,
20	44.5	45.4	852	12	US-09-971-101A-4	Sequence 4, Appli
21	44	44.9	212	9	US-09-925-300-1577	Sequence 1577, Ap
22	44	44.9	251	12	US-10-425-114-43396	Sequence 43396, A
23	43.5	44.4	1367	15	US-10-320-797-3355	Sequence 3355, Ap
24	43	43.9	52	9	US-09-864-761-39967	Sequence 39967, A
25	43	43.9	300	15	US-10-369-493-12997	Sequence 12997, A
26	43	43.9	387	16	US-10-437-963-118548	Sequence 118548,
27	43	43.9	448	12	US-10-425-114-54044	Sequence 54044, A
28	43	43.9	712	15	US-10-369-493-3977	Sequence 3977, Ap
29	43	43.9	1165	16	US-10-408-765A-1392	Sequence 1392, Ap
30	43	43.9	1212	16	US-10-618-581-5	Sequence 5, Appli
31	42	42.9	51	12	US-10-424-599-247293	Sequence 247293,
32	42	42.9	264	14	US-10-017-161-706	Sequence 706, App
33	42	42.9	340	16	US-10-389-566-1425	Sequence 1425, Ap
34	42	42.9	340	16	US-10-389-566-1831	Sequence 1831, Ap
35	42	42.9	364	12	US-10-425-114-41207	Sequence 41207, A
36	42	42.9	399	9	US-09-764-870-409	Sequence 409, App
37	42	42.9	399	14	US-10-125-540-409	Sequence 409, App
38	42	42.9	421	16	US-10-437-963-133514	Sequence 133514,
39	42	42.9	457	11	US-09-826-509-579	Sequence 579, App
40	42	42.9	457	14	US-10-225-567A-469	Sequence 469, App
41	42	42.9	457	15	US-10-292-798-618	Sequence 618, App
42	42	42.9	495	15	US-10-295-027-875	Sequence 875, App
43	42	42.9	506	9	US-09-814-986-8	Sequence 8, Appli
44	42	42.9	506	10	US-09-782-390-4	Sequence 4, Appli
45	42	42.9	518	16	US-10-437-963-195842	Sequence 195842,

ALIGNMENTS

RESULT 1

US-10-281-652-3

; Sequence 3, Application US/10281652

; Publication No. US20030091606A1

; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John

```
; APPLICANT: HUGHES, Thomas K.
; APPLICANT: BOLDOGH, Istvan
; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND
; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS
; FILE REFERENCE: 265.00220101
; CURRENT APPLICATION NUMBER: US/10/281,652
; CURRENT FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US/09/641,803
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,310
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: peptide
US-10-281-652-3
```

```
Query Match          100.0%; Score 98; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 DQPPDVEKPDLPFQVQS 18
          |||||
Db      1 DQPPDVEKPDLPFQVQS 18
```

RESULT 2

US-10-282-122A-56851

```
; Sequence 56851, Application US/10282122A
```

```
; Publication No. US20040029129A1
```

GENERAL INFORMATION:

```
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
; FILE REFERENCE: ELITRA.034A
```

```
; CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
```

```
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56851
;   LENGTH: 1047
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-10-282-122A-56851
```

```
Query Match          52.0%;   Score 51;   DB 12;   Length 1047;
Best Local Similarity 47.1%;   Pred. No. 92;
Matches      8;   Conservative    6;   Mismatches    3;   Indels      0;   Gaps      0;
```

```
Qy      1 DQPPDVEKPDLPFQVQ 17
        ||| |::||: : ||::
Db      256 DQPVDLQKPETKQFQLK 272
```

RESULT 3

```
US-10-210-130-118
; Sequence 118, Application US/10210130
; Publication No. US20040014053A1
; GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Patturajan, Meera
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Berghs, Constance
; APPLICANT: Zhong, Mei
; APPLICANT: Casman, Stacie J.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Smithson, Glennda
; APPLICANT: Ji, Weizhen
; APPLICANT: Gorman, Linda
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Leite, Mario W.
```

; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Anderson, David W.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Khramtsov, Nikolai V.
; APPLICANT: Ort, Tatiana
; APPLICANT: Ellerman, Karen
; APPLICANT: Rastelli, Luca
; APPLICANT: Agee, Michele L.
; APPLICANT: Chaudhuri, Amitabha
; APPLICANT: Chant, John S.
; APPLICANT: DiPippo, Vincent A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Giot, Loic
; APPLICANT: Ooi, Chean Eng
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Hjalt, Tord
; APPLICANT: Liu, Xiaohong
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Catterton, Elina
; APPLICANT: Shenoy, Suresh G.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-416C (Cura-716 SMT)
; CURRENT APPLICATION NUMBER: US/10/210,130
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 60/309,501
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/316,508
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/354,655
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 60/310,291
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/383,887
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/323,936
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: 60/381,039
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 118
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-210-130-118

Query Match 49.0%; Score 48; DB 15; Length 377;
Best Local Similarity 72.7%; Pred. No. 84;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
|||||:| ||
Db 361 PPDVEQPQTQP 371

RESULT 4

US-10-250-613-1

; Sequence 1, Application US/10250613

; Publication No. US20040096828A1

; GENERAL INFORMATION:

; APPLICANT: LU, Dyung Aina M.; BAUGHN, Mariah R.;

; APPLICANT: YAO, Monique G.; DING, Li;

; APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;

; APPLICANT: TANG, Y. Tom; WARREN, Bridget A.;

; APPLICANT: DUGGAN, Brendan M.; XU, Yuming;

; APPLICANT: CHAWLA, Narinder K.; GRIFFIN, Jennifer A.;

; APPLICANT: STEWART, Elizabeth A.; GANDHI, Ameena R.;

; APPLICANT: KHAN, Farrah A.; THANGAVELU, Kavitha;

; APPLICANT: ISON, Craig H.; AZIMZAI, Yalda;

; APPLICANT: HAFALIA, April J.A.; GIETZEN, Kimberly J.;

; APPLICANT: LAL, Preeti G.; SANJANWALA, Madhusudan M.;

; APPLICANT: ELLIOTT, Vicki S.

; TITLE OF INVENTION: CYTOSKELETAL-ASSOCIATED PROTEINS

; FILE REFERENCE: PF-0878 USN

; CURRENT APPLICATION NUMBER: US/10/250,613

; CURRENT FILING DATE: 2003-07-02

; PRIOR APPLICATION NUMBER: PCT/US02/00178

; PRIOR FILING DATE: 2002-01-04

; PRIOR APPLICATION NUMBER: US 60/260,085

; PRIOR FILING DATE: 2001-01-04

; PRIOR APPLICATION NUMBER: US 60/268,554

; PRIOR FILING DATE: 2001-02-13

; PRIOR APPLICATION NUMBER: US 60/269,111

; PRIOR FILING DATE: 2001-02-14

; PRIOR APPLICATION NUMBER: US 60/271,211

; PRIOR FILING DATE: 2001-02-23

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: PERL Program

; SEQ ID NO 1

; LENGTH: 377

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No: 5566074CD1

US-10-250-613-1

Query Match 49.0%; Score 48; DB 16; Length 377;
Best Local Similarity 72.7%; Pred. No. 84;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13

Db |||||:| ||
361 PPDVEQPQTQP 371

RESULT 5
US-09-815-242-10746
; Sequence 10746, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10746
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10746

Query Match 49.0%; Score 48; DB 9; Length 541;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQV 16
 || | || ||| :|
Db 274 DQQPGKEKWDLPMEV 289

RESULT 6
US-10-437-963-204178
; Sequence 204178, Application US/10437963
; Publication No. US20040123343A1

```

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 204178
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_99291C.1.pep
US-10-437-963-204178

```

```

Query Match          48.0%; Score 47; DB 16; Length 162;
Best Local Similarity 100.0%; Pred. No. 47;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy          3 PPDVEKPD 10
             |||||
Db          15 PPDVEKPD 22

```

```

RESULT 7
US-10-437-963-200861
; Sequence 200861, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 200861
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Oryza sativa

```

; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_96290C.1.pep
US-10-437-963-200861

Query Match 46.9%; Score 46; DB 16; Length 212;
Best Local Similarity 52.9%; Pred. No. 88;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 2 QPPDVEKPDLPFQVQS 18
:|| ||: ||| ||
Db 36 RPPLRSKPEALPFQAQS 52

RESULT 8

US-10-425-114-52231
; Sequence 52231, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52231
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700894460_FLI.pep
US-10-425-114-52231

Query Match 46.9%; Score 46; DB 12; Length 313;
Best Local Similarity 53.3%; Pred. No. 1.3e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 QPPDVEKPDLPFQV 16
|| ::||: || ||
Db 151 QPVELEEPNQQLQV 165

RESULT 9

US-10-424-599-233789
; Sequence 233789, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 233789
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53136C.1.pep
US-10-424-599-233789

Query Match 46.9%; Score 46; DB 12; Length 336;
Best Local Similarity 53.3%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 QPPDVEKPDLPFFQV 16
|| ::|:|: || ||
Db 149 QPVELEEPNQQLQV 163

RESULT 10

US-10-425-114-39945

; Sequence 39945, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 39945

; LENGTH: 337
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:

; OTHER INFORMATION: Clone ID: 700904208_FLI.pep
US-10-425-114-39945

Query Match 46.9%; Score 46; DB 12; Length 337;
Best Local Similarity 53.3%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 QPPDVEKPDLPFFQV 16
|| ::|:|: || ||
Db 151 QPVELEEPNQQLQV 165

RESULT 11

US-09-880-149-13

; Sequence 13, Application US/09880149
 ; Patent No. US20020146843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kenten, John
 ; APPLICANT: Roberts, Steven
 ; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUCARYOTIC ORGANISMS
 ; FILE REFERENCE: 2757-5
 ; CURRENT APPLICATION NUMBER: US/09/880,149
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: 09/406,781
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: 60/119,851
 ; PRIOR FILING DATE: 1999-02-12
 ; NUMBER OF SEQ ID NOS: 67
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 26
 ; TYPE: PRT
 ; ORGANISM: Unknown Organism
 ; FEATURE:
 ; OTHER INFORMATION: Description of Unknown Organism: PEST example
 ; OTHER INFORMATION: sequence
 US-09-880-149-13

Query Match 45.9%; Score 45; DB 9; Length 26;
 Best Local Similarity 57.1%; Pred. No. 13;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPQPFQV 16
 || ||:||: | |
 Db 2 PPGVEEPDVGPLPV 15

RESULT 12

US-09-880-132-13

; Sequence 13, Application US/09880132
 ; Patent No. US20020173049A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kenten, John
 ; APPLICANT: Roberts, Steven
 ; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUCARYOTIC ORGANISMS
 ; FILE REFERENCE: 2757-6
 ; CURRENT APPLICATION NUMBER: US/09/880,132
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: 09/406,781
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: 60/119,851
 ; PRIOR FILING DATE: 1999-02-12
 ; NUMBER OF SEQ ID NOS: 67
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 26
 ; TYPE: PRT

; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: PEST example
; OTHER INFORMATION: sequence
US-09-880-132-13

Query Match 45.9%; Score 45; DB 9; Length 26;
Best Local Similarity 57.1%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPQPFQV 16
|| ||: ||: | |
Db 2 PPGVEEPDVGPLPV 15

RESULT 13

US-10-345-281-13
; Sequence 13, Application US/10345281
; Publication No. US20030153727A1
; GENERAL INFORMATION:
; APPLICANT: Kenten, John
; APPLICANT: Roberts, Steven
; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUKARYOTIC ORGANISMS
; FILE REFERENCE: 2757-6
; CURRENT APPLICATION NUMBER: US/10/345,281
; CURRENT FILING DATE: 2003-01-16
; PRIOR APPLICATION NUMBER: US/09/880,132
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 09/406,781
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: 60/119,851
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: PEST example
; OTHER INFORMATION: sequence
US-10-345-281-13

Query Match 45.9%; Score 45; DB 14; Length 26;
Best Local Similarity 57.1%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPQPFQV 16
|| ||: ||: | |
Db 2 PPGVEEPDVGPLPV 15

RESULT 14

US-10-424-599-228647
; Sequence 228647, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
 With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 228647
 ; LENGTH: 159
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_48496C.1.pep
 US-10-424-599-228647

Query Match 45.9%; Score 45; DB 12; Length 159;
 Best Local Similarity 58.3%; Pred. No. 90;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 5 DVEKPDLPFQV 16
 ||:||||:| |:
 Db 145 DVKKPDVKPVQI 156

RESULT 15

US-09-925-302-766
 ; Sequence 766, Application US/09925302
 ; Patent No. US20020044941A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA104
 ; CURRENT APPLICATION NUMBER: US/09/925,302
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05918
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 896
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 766
 ; LENGTH: 848
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (2)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (8)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-302-766

Query Match 45.9%; Score 45; DB 9; Length 848;
Best Local Similarity 41.2%; Pred. No. 5.4e+02;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
||||: :|| :: |:
Db 569 DQPPEAKKPKIKVVNVE 585

Search completed: August 24, 2004, 16:41:16
Job time : 68.1493 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:23:00 ; Search time 55.6119 Seconds
(without alignments)
102.124 Million cell updates/sec

Title: US-09-641-801-3
Perfect score: 98
Sequence: 1 DQPPDVEKPDLPQPFQVQS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	ID	Description
<hr/>					

1	55	56.1	483	16	Q9A5S4	Q9a5s4 caulobacter
2	51	52.0	182	16	Q9X6Z0	Q9x6z0 bordetella
3	51	52.0	182	16	Q7WGH6	Q7wgh6 bordetella
4	51	52.0	182	16	Q7W511	Q7w511 bordetella
5	51	52.0	392	2	O88168	O88168 enterococcu
6	51	52.0	1047	16	Q832P0	Q832p0 enterococcu
7	48	49.0	173	4	Q7Z302	Q7z302 homo sapien
8	48	49.0	377	4	Q8N426	Q8n426 homo sapien
9	48	49.0	541	16	Q835M9	Q835m9 enterococcu
10	47	48.0	404	3	Q99118	Q99118 ustilago ma
11	47	48.0	516	10	Q9C8M8	Q9c8m8 arabidopsis
12	47	48.0	719	10	Q9C8M9	Q9c8m9 arabidopsis
13	46.5	47.4	126	10	Q84R17	Q84r17 arabidopsis
14	46	46.9	134	16	Q9X8L8	Q9x8l8 streptomyce
15	46	46.9	137	11	Q9D2D5	Q9d2d5 mus musculu
16	46	46.9	141	11	Q9QW43	Q9qw43 mus sp. ame
17	46	46.9	154	11	Q61293	Q61293 mus musculu
18	46	46.9	195	11	Q63640	Q63640 rattus norv
19	46	46.9	199	16	Q8UEB8	Q8ueb8 agrobacteri
20	46	46.9	210	11	P70592	P70592 rattus norv
21	46	46.9	219	11	Q62945	Q62945 rattus norv
22	46	46.9	499	5	Q9V539	Q9v539 drosophila
23	46	46.9	627	13	Q7ZXG8	Q7zxcg8 xenopus lae
24	46	46.9	728	16	Q9RYM8	Q9rym8 deinococcus
25	46	46.9	908	5	Q8MMF4	Q8mmf4 drosophila
26	45	45.9	194	11	O88326	O88326 mesocricetu
27	45	45.9	345	11	Q8C6H1	Q8c6h1 mus musculu
28	45	45.9	369	13	O73737	O73737 gallus gall
29	45	45.9	391	11	Q8BHT7	Q8bht7 mus musculu
30	45	45.9	858	11	Q8VCW6	Q8vcw6 mus musculu
31	45	45.9	858	11	Q8C430	Q8c430 mus musculu
32	45	45.9	1572	16	Q8G7T8	Q8g7t8 bifidobacte
33	45	45.9	6994	5	Q17343	Q17343 caenorhabdi
34	45	45.9	6994	5	Q17490	Q17490 caenorhabdi
35	44.5	45.4	380	2	O88128	O88128 vibrio para
36	44.5	45.4	502	5	Q8IR21	Q8ir21 drosophila
37	44.5	45.4	739	11	Q7TPY7	Q7tpy7 mus musculu
38	44.5	45.4	852	11	Q9QUG2	Q9qug2 mus musculu
39	44.5	45.4	1520	5	Q9VXJ7	Q9vxj7 drosophila
40	44	44.9	175	10	Q9ZQB6	Q9zqb6 arabidopsis
41	44	44.9	192	4	Q9H9L7	Q9h9l7 homo sapien
42	44	44.9	216	10	Q8L814	Q8l814 arabidopsis
43	44	44.9	230	10	Q8LGD2	Q8lgd2 arabidopsis
44	44	44.9	277	10	Q8RXV8	Q8rxv8 arabidopsis
45	44	44.9	328	11	Q8C643	Q8c643 mus musculu

ALIGNMENTS

RESULT 1

Q9A5S4

ID Q9A5S4 PRELIMINARY; PRT; 483 AA.

AC Q9A5S4;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Methylmalonyl-CoA mutase, beta subunit.
 GN CC2373.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=211173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005906; AAK24344.1; -.
 DR PIR; D87543; D87543.
 DR HSSP; P11652; 1REQ.
 DR TIGR; CC2373; -.
 DR GO; GO:0004494; F:methylmalonyl-CoA mutase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR006099; MMCoA_mutase.
 DR Pfam; PF01642; MM_CoA_mutase; 1.
 KW Complete proteome.
 SQ SEQUENCE 483 AA; 50032 MW; 194F84D33268D6D5 CRC64;

Query Match 56.1%; Score 55; DB 16; Length 483;
 Best Local Similarity 58.8%; Pred. No. 2.7;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DQPPDVEKPDLPFQVQ 17
 |:|:|:| | | | | |
 Db 439 DKPPEVETPDSSAFVQ 455

RESULT 2
 Q9X6Z0

ID Q9X6Z0 PRELIMINARY; PRT; 182 AA.
 AC Q9X6Z0;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Outer membrane lipoprotein.
 GN OMLA OR BP2508.
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TohamaI;
 RA Pradel E.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
 RA Rabbिनowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,
 RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; AJ238308; CAB41013.1; -.
 DR EMBL; BX640418; CAE42780.1; -.
 DR InterPro; IPR000437; Prok_lipoprot_S.
 DR InterPro; IPR007450; SmpA_OmlA.
 DR Pfam; PF04355; SmpA_OmlA; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 182 AA; 20489 MW; 73F6DB9B171AD791 CRC64;

Query Match 52.0%; Score 51; DB 16; Length 182;
 Best Local Similarity 72.7%; Pred. No. 4;
 Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 7 EKPDLPFQVQ 17
 |:|||||:::
 Db 114 EQPDLPFQIE 124

RESULT 3

Q7WGH6

ID Q7WGH6 PRELIMINARY; PRT; 182 AA.
 AC Q7WGH6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Outer membrane lipoprotein.
 GN OMLA OR BB3943.
 OS *Bordetella bronchiseptica* (*Alcaligenes bronchisepticus*).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; *Bordetella*.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,

RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
 RA Rabbinoiwitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640449; CAE34306.1; -.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 182 AA; 20504 MW; DCF6DB9B17142113 CRC64;

Query Match 52.0%; Score 51; DB 16; Length 182;
 Best Local Similarity 72.7%; Pred. No. 4;
 Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 7 EKPDLQPFQVQ 17
 |:|||||:::
 Db 114 EQPDLQPFQIE 124

RESULT 4

Q7W511

ID Q7W511 PRELIMINARY; PRT; 182 AA.
 AC Q7W511;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Outer membrane lipoprotein.
 GN OMLA OR BPP3495.
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=12822 / ATCC BAA-587;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
 RA Rabbinoiwitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640433; CAE38779.1; -.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 182 AA; 20490 MW; 73F6DB9B1714377F CRC64;

Query Match 52.0%; Score 51; DB 16; Length 182;
 Best Local Similarity 72.7%; Pred. No. 4;

Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 7 EKPDLQPFQVQ 17
 |:|||||:|:
 Db 114 EQPDLQPFQIE 124

RESULT 5

O88168

ID O88168 PRELIMINARY; PRT; 392 AA.
 AC O88168;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Orfdel4.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OG1RF;
 RX MEDLINE=98380380; PubMed=9712783;
 RA Xu Y., Murray B.E., Weinstock G.M.;
 RT "A cluster of genes involved in polysaccharide biosynthesis from
 RT Enterococcus faecalis OG1RF."
 RL Infect. Immun. 66:4313-4323(1998).
 DR EMBL; AF071085; AAC35928.1; -.
 DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
 DR InterPro; IPR001601; Methyltransf.
 DR InterPro; IPR000051; SAM bind.
 SQ SEQUENCE 392 AA; 44996 MW; 687A988FC2078CF6 CRC64;

Query Match 52.0%; Score 51; DB 2; Length 392;
 Best Local Similarity 47.1%; Pred. No. 9.1;
 Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLQPFQVQ 17
 ||| |::||: : ||::
 Db 256 DQPVDLQKPETKQFQLK 272

RESULT 6

Q832P0

ID Q832P0 PRELIMINARY; PRT; 1047 AA.
 AC Q832P0;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Glycosyl transferase, group 2 family protein.
 GN EF2181.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V583 / ATCC 700802;
 RX MEDLINE=22550857; PubMed=12663927;

RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
 RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
 RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
 RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
 RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
 RT "Role of mobile DNA in the evolution of vancomycin-resistant
 RT Enterococcus faecalis.";
 RL Science 299:2071-2074(2003).
 DR EMBL; AE016953; AA081913.1; -.
 DR TIGR; EF2181; -.
 DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR001173; Glyco_trans_2.
 DR InterPro; IPR001601; Methyltransf.
 DR InterPro; IPR000051; SAM_bind.
 DR Pfam; PF00535; Glycos_transf_2; 1.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 1047 AA; 119728 MW; 621F8B792F814E36 CRC64;

Query Match 52.0%; Score 51; DB 16; Length 1047;
 Best Local Similarity 47.1%; Pred. No. 26;
 Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPQFQVQ 17
 ||| |::||: : ||::
 Db 256 DQPVDLQKPETKQFQLK 272

RESULT 7

Q7Z302

ID Q7Z302 PRELIMINARY; PRT; 173 AA.
 AC Q7Z302;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein DKFZp686I16132 (Fragment).
 GN DKFZP686I16132.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Human uterus;
 RA Lauber J., Bahr A., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
 RA Han M., Wiemann S.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BX538316; CAD98091.1; -.
 KW Hypothetical protein.
 FT NON_TER 1 1
 SQ SEQUENCE 173 AA; 19927 MW; 0774F47B1D71E344 CRC64;

Query Match 49.0%; Score 48; DB 4; Length 173;
 Best Local Similarity 72.7%; Pred. No. 11;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
|||||:| ||
Db 157 PPDVEQPQTQP 167

RESULT 8

Q8N426

ID Q8N426 PRELIMINARY; PRT; 377 AA.
AC Q8N426;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC036819; AAH36819.1; -.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
KW Hypothetical protein.
SQ SEQUENCE 377 AA; 43212 MW; 7A13E2C28E1AD6EA CRC64;

Query Match 49.0%; Score 48; DB 4; Length 377;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
|||||:| ||
Db 361 PPDVEQPQTQP 371

RESULT 9

Q835M9

ID Q835M9 PRELIMINARY; PRT; 541 AA.
AC Q835M9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glucan 1,6-alpha-glucosidase, putative.
GN EF1348.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,

RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
 RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
 RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
 RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
 RT "Role of mobile DNA in the evolution of vancomycin-resistant
 RT *Enterococcus faecalis*.";
 RL Science 299:2071-2074(2003).
 DR EMBL; AE016951; AA081139.1; -.
 DR TIGR; EF1348; -.
 DR GO; GO:0004556; F:alpha-amylase activity; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR006047; Alpha_amyl_cat.
 DR InterPro; IPR006589; Alp_amyl_cat_sub.
 DR Pfam; PF00128; alpha-amylase; 1.
 DR SMART; SM00642; Amy; 1.
 KW Complete proteome.
 SQ SEQUENCE 541 AA; 62718 MW; ED0DB68653A7DC98 CRC64;

Query Match 49.0%; Score 48; DB 16; Length 541;
 Best Local Similarity 62.5%; Pred. No. 37;
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLQPFQV 16
 || | || ||| :|
 Db 274 DQQPGKEKWDLPMEV 289

RESULT 10

Q99118

ID Q99118 PRELIMINARY; PRT; 404 AA.
 AC Q99118;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE BW4 protein (Fragment).
 GN BW4.
 OS *Ustilago maydis* (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; *Ustilago*.
 OX NCBI_TaxID=5270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RK138;
 RX MEDLINE=92154679; PubMed=1739973;
 RA Gillissen B., Bergemann J., Sandmann C., Schroeer B., Boelker M.,
 RA Kahmann R.;
 RT "a two-component regulatory system for self/non-self recognition in
 RT *ustilago maydis*.";
 RL Cell 68:647-657(1992).
 DR EMBL; M84181; AAA34223.1; -.
 DR PIR; D42094; D42094.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 FT NON_TER 404 404
 SQ SEQUENCE 404 AA; 45439 MW; 4B2C71857AC82910 CRC64;

Query Match 48.0%; Score 47; DB 3; Length 404;
 Best Local Similarity 47.1%; Pred. No. 39;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 QPPDVEKPDLPFFQVQS 18
 :| | :||| ||: |
 Db 198 EPTDSTQPDLSFRRSES 214

RESULT 11

Q9C8M8

ID Q9C8M8 PRELIMINARY; PRT; 516 AA.
 AC Q9C8M8;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Leucine-rich repeat transmembrane protein kinase 1, putative, 10414-
 DE 7611.
 GN F22G10.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana."
 RL Nature 408:816-820(2000).
 DR EMBL; AC024260; AAG51973.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004672; F:protein kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007090; LRR_plant.
 DR InterPro; IPR000719; Prot_kinase.
 DR Pfam; PF00560; LRR; 5.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Kinase; Transferase; Transmembrane.
 SQ SEQUENCE 516 AA; 55793 MW; DCB17185AFC935C1 CRC64;

Query Match 48.0%; Score 47; DB 10; Length 516;
 Best Local Similarity 56.2%; Pred. No. 51;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFFQVQS 18
 | |:| | ||| : |
 Db 310 PMDIEKTDNQPF TLAS 325

RESULT 12

Q9C8M9

ID Q9C8M9 PRELIMINARY; PRT; 719 AA.
 AC Q9C8M9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Leucine-rich repeat transmembrane protein kinase 1, putative, 10414-
 DE 6710.
 GN F22G10.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

RT thaliana.";
 RL Nature 408:816-820(2000).
 DR EMBL; AC024260; AAG51974.1; -.
 DR PIR; F96577; F96577.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004672; F:protein kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007090; LRR_plant.
 DR InterPro; IPR000719; Prot_kinase.
 DR Pfam; PF00560; LRR; 5.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Kinase; Transferase; Transmembrane.
 SQ SEQUENCE 719 AA; 78084 MW; 88189A8C71B64412 CRC64;

Query Match 48.0%; Score 47; DB 10; Length 719;
 Best Local Similarity 56.2%; Pred. No. 72;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFFQVQS 18
 | | : | | | : |
 Db 324 PMDIEKTDNQPF TLAS 339

RESULT 13

Q84R17

ID Q84R17 PRELIMINARY; PRT; 126 AA.
 AC Q84R17;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein At4g24275.
 GN AT4G24275.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Arabidopsis Full Length cDNA Clones.";
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BT006169; AAP04153.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 126 AA; 14395 MW; 01B6C544EEDF614D CRC64;

Query Match 47.4%; Score 46.5; DB 10; Length 126;
Best Local Similarity 40.0%; Pred. No. 14;
Matches 8; Conservative 6; Mismatches 3; Indels 3; Gaps 1;

QY 1 DQP---PDVEKPDLPFFQVQ 17
| | | ::|||:| :::
Db 17 DHPVVDPQIQKPDLEPAEMK 36

RESULT 14

Q9X8L8

ID Q9X8L8 PRELIMINARY; PRT; 134 AA.
AC Q9X8L8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical proline-rich protein.
GN SCO3354 OR SCE94.05.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabbिनowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";

RL Nature 417:141-147(2002).
DR EMBL; AL939116; CAB40854.1; -.
DR PIR; T36365; T36365.
KW Complete proteome.
SQ SEQUENCE 134 AA; 13690 MW; C2526A91D4C47806 CRC64;

Query Match 46.9%; Score 46; DB 16; Length 134;
Best Local Similarity 57.1%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPF 14
| | | | | : | |
Db 34 DPPPDSPPPDPEPF 47

RESULT 15

Q9D2D5

ID Q9D2D5 PRELIMINARY; PRT; 137 AA.

AC Q9D2D5;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE 4932443L11Rik protein.

GN 4932443L11RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL; AK019856; BAB31884.1; -.

DR MGD; MGI:1926149; 4932443L11Rik.

SQ SEQUENCE 137 AA; 15013 MW; 22FE9AF363DC21D2 CRC64;

Query Match 46.9%; Score 46; DB 11; Length 137;

Best Local Similarity 58.3%; Pred. No. 18;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPF 14
||:|:| | |:
Db 66 PPEVEQPSLPPY 77

Search completed: August 24, 2004, 15:50:31
Job time : 60.6119 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:57:04 ; Search time 9.67164 Seconds
(without alignments)
96.908 Million cell updates/sec

Title: US-09-641-801-3
Perfect score: 98
Sequence: 1 DQPPDVEKPDLPFQVQS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	49.0	461	1	DCOR_BOVIN	P27117 bos taurus
2	46	46.9	196	1	AMEX_MOUSE	P45559 mus musculu
3	45	45.9	391	1	Y258_HUMAN	Q92546 homo sapien
4	45	45.9	858	1	H105_CRIGR	Q60446 cricetulus
5	45	45.9	858	1	H105_HUMAN	Q92598 homo sapien
6	45	45.9	858	1	H105_MOUSE	Q61699 mus musculu
7	44	44.9	1070	1	DNL1_XENLA	P51892 xenopus lae
8	44	44.9	1230	1	SAH1_MOUSE	P59808 mus musculu
9	43	43.9	699	1	NP14_HUMAN	Q14978 homo sapien
10	43	43.9	1165	1	Z407_HUMAN	Q9c0g0 homo sapien
11	43	43.9	1462	1	DPOA_HUMAN	P09884 homo sapien
12	42.5	43.4	2774	1	MAPA_RAT	P34926 rattus norv
13	42	42.9	254	1	SLBP_XENLA	P79943 xenopus lae
14	42	42.9	304	1	IG1R_PIG	Q29000 sus scrofa
15	42	42.9	372	1	HXA2_MOUSE	P31245 mus musculu
16	42	42.9	457	1	VIPR_HUMAN	P32241 homo sapien
17	42	42.9	506	1	TUB_HUMAN	P50607 homo sapien

18	42	42.9	640	1	IG1R_BOVIN	Q05688	bos taurus
19	41	41.8	148	1	YLL2_EBV	P03199	epstein-bar
20	41	41.8	148	1	YLL2_EBVA8	Q07285	epstein-bar
21	41	41.8	189	1	CD3E_MOUSE	P22646	mus musculu
22	41	41.8	192	1	CD3E_SHEEP	P29328	ovis aries
23	41	41.8	202	1	CD3E_CANFA	P27597	canis famil
24	41	41.8	207	1	CD3E_HUMAN	P07766	homo sapien
25	41	41.8	498	1	IRF5_HUMAN	Q13568	homo sapien
26	41	41.8	704	1	VPS1_YEAST	P21576	saccharomyc
27	41	41.8	807	1	COLA_VIBVU	Q8d4y9	vibrio vuln
28	41	41.8	851	1	DYN1_RAT	P21575	rattus norv
29	41	41.8	864	1	DYN1_HUMAN	Q05193	homo sapien
30	41	41.8	867	1	DYN1_MOUSE	P39053	mus musculu
31	41	41.8	916	1	DNL1_MOUSE	P37913	mus musculu
32	41	41.8	918	1	DNL1_RAT	Q9jhy8	rattus norv
33	41	41.8	919	1	DNL1_HUMAN	P18858	homo sapien
34	41	41.8	2351	1	FA8_HUMAN	P00451	homo sapien
35	41	41.8	5035	1	RYR1_PIG	P16960	sus scrofa
36	40.5	41.3	815	1	CC53_YEAST	Q12018	saccharomyc
37	40.5	41.3	2805	1	MAPA_HUMAN	P78559	homo sapien
38	40	40.8	227	1	RS3_METKA	Q8tx35	methanopyru
39	40	40.8	228	1	HS74_LEIMA	P12077	leishmania
40	40	40.8	300	1	NUSG_STRCO	P36266	streptomyce
41	40	40.8	327	1	MOXR_RAT	Q9es58	rattus norv
42	40	40.8	343	1	UL14_HCMVA	P16756	human cytom
43	40	40.8	358	1	PIT1_ONCMY	Q08478	oncorhynchu
44	40	40.8	365	1	PIT1_ONCKE	Q91169	oncorhynchu
45	40	40.8	368	1	TGF4_MOUSE	Q64280	mus musculu

ALIGNMENTS

RESULT 1

DCOR_BOVIN

ID DCOR_BOVIN STANDARD; PRT; 461 AA.

AC P27117;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ornithine decarboxylase (EC 4.1.1.17) (ODC).

GN ODC1 OR ODC.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Holstein; TISSUE=Liver;

RX MEDLINE=95293216; PubMed=7774801;

RA Yao J., Zadworny D., Kuhnlein U., Hayes J.F.;

RT "Molecular cloning of a bovine ornithine decarboxylase cDNA and its
RT use in the detection of restriction fragment length polymorphisms in
RT Holsteins.";

RL Genome 38:325-331(1995).

CC -!- CATALYTIC ACTIVITY: L-ornithine = putrescine + CO(2).

CC -!- COFACTOR: Pyridoxal phosphate.
 CC -!- PATHWAY: Polyamine biosynthesis; first (rate-limiting) step.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF ORNITHINE, DAP, AND ARGININE
 CC DECARBOXYLASES.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M92441; AAA92339.1; -.
 DR EMBL; U36394; AAA79849.1; -.
 DR EMBL; U18531; AAA86696.1; -.
 DR HSSP; P11926; 1D7K.
 DR InterPro; IPR000183; Decarboxylase2.
 DR InterPro; IPR009006; Racem_decarbox_C.
 DR Pfam; PF02784; Orn_Arg_deC_N; 1.
 DR Pfam; PF00278; Orn_DAP_Arg_deC; 1.
 DR PRINTS; PR01179; ODADCRBXLASE.
 DR PROSITE; PS00878; ODR_DC_2_1; 1.
 DR PROSITE; PS00879; ODR_DC_2_2; 1.
 KW Lyase; Decarboxylase; Pyridoxal phosphate; Polyamine biosynthesis;
 KW Phosphorylation.
 FT BINDING 69 69 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 FT ACT_SITE 360 360 BY SIMILARITY.
 FT MOD_RES 303 303 PHOSPHORYLATION (BY CK2)
 FT (BY SIMILARITY).
 SQ SEQUENCE 461 AA; 51345 MW; 4E609B643E3B68FA CRC64;

Query Match 49.0%; Score 48; DB 1; Length 461;
 Best Local Similarity 56.2%; Pred. No. 6.5;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQV 16
 | || ||: ||: | |
 Db 424 DFPPGVVEPDVGPLPV 439

RESULT 2

AMEX_MOUSE

ID AMEX_MOUSE STANDARD; PRT; 196 AA.
 AC P45559;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Amelogenin, X isoform precursor (Leucine-rich amelogenin peptide)
 DE (LRAP).
 GN AMELX OR AMEL.
 OS Mus musculus (Mouse), and
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090, 10116;

RN [1]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RC SPECIES=Mouse;
 RX MEDLINE=85251692; PubMed=4015654;
 RA Snead M.L., Lau E.C., Zeichner-David M., Fincham A.G., Woo S.L.,
 RA Slavkin H.C.;
 RT "DNA sequence for cloned cDNA for murine amelogenin reveal the amino
 RT acid sequence for enamel-specific protein.";
 RL Biochem. Biophys. Res. Commun. 129:812-818(1985).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC SPECIES=Mouse; STRAIN=ICR;
 RX MEDLINE=93075222; PubMed=1445358;
 RA Lau E.C., Simmer J.P., Bringas P. Jr., Hsu D.D.J., Hu C.C.,
 RA Zeichner-David M., Thiemann F., Snead M.L., Slavkin H.C.,
 RA Fincham A.G.;
 RT "Alternative splicing of the mouse amelogenin primary RNA transcript
 RT contributes to amelogenin heterogeneity.";
 RL Biochem. Biophys. Res. Commun. 188:1253-1260(1992).
 RN [3]
 RP REVISION TO 4.
 RC SPECIES=Mouse; STRAIN=ICR;
 RA Oida S., Iimura T., Arai N., Takeda K., Maruoka Y., Terashima T.,
 RA Shimokawa H., Sasaki S.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Rat; STRAIN=Wistar; TISSUE=Enamel organ;
 RX MEDLINE=94128126; PubMed=8297387;
 RA Bonass W.A., Robinson P.A., Kirkham J., Shore R.C., Robinson C.;
 RT "Molecular cloning and DNA sequence of rat amelogenin and a
 RT comparative analysis of mammalian amelogenin protein sequence
 RT divergence.";
 RL Biochem. Biophys. Res. Commun. 198:755-763(1994).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC SPECIES=Rat; STRAIN=Wistar; TISSUE=Enamel organ;
 RX MEDLINE=95035099; PubMed=7948026;
 RA Bonass W.A., Kirkham J., Brookes S.J., Shore R.C., Robinson C.;
 RT "Isolation and characterisation of an alternatively-spliced rat
 RT amelogenin cDNA: LRAP -- a highly conserved, functional alternatively-
 RT spliced amelogenin?";
 RL Biochim. Biophys. Acta 1219:690-692(1994).
 CC -!- FUNCTION: PLAYS A ROLE IN THE BIOMINERALIZATION OF TEETH. SEEMS TO
 CC REGULATE THE FORMATION OF CRYSTALLITES DURING THE SECRETORY STAGE
 CC OF TOOTH ENAMEL DEVELOPMENT. THOUGHT TO PLAY A MAJOR ROLE IN THE
 CC STRUCTURAL ORGANIZATION AND MINERALIZATION OF DEVELOPING ENAMEL.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=P45559-1; Sequence=Displayed;
 CC Name=2; Synonyms=LRAP;
 CC IsoId=P45559-2; Sequence=VSP_000230;
 CC Name=3;
 CC IsoId=P45559-3; Sequence=VSP_000231;

CC -!- PTM: Several forms are produced by carboxy-terminal processing.
 CC -!- SIMILARITY: Belongs to the amelogenin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D31768; BAA06546.1; -.
 DR EMBL; D31769; BAA06547.1; -.
 DR EMBL; U01245; AAA20491.1; -.
 DR EMBL; U07054; AAA61964.1; -.
 DR PIR; JC2391; JC2391.
 DR MGD; MGI:88005; Amelx.
 DR GO; GO:0005578; C:extracellular matrix; ISS.
 DR GO; GO:0030345; F:structural constituent of tooth enamel; ISS.
 DR GO; GO:0030282; P:bone mineralization; ISS.
 DR GO; GO:0042476; P:odontogenesis; ISS.
 DR InterPro; IPR004116; Amelogenin.
 DR Pfam; PF02948; Amelogenin; 1.
 DR PRINTS; PR01757; AMELOGENIN.
 KW Biomineralization; Extracellular matrix; Phosphorylation; Repeat;
 KW Signal; Alternative splicing.
 FT SIGNAL 1 16 BY SIMILARITY.
 FT CHAIN 17 196 AMELOGENIN, X ISOFORM.
 FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
 FT VARSPLIC 50 170 Missing (in isoform 2).
 FT /FTId=VSP_000230.
 FT VARSPLIC 50 73 Missing (in isoform 3).
 FT /FTId=VSP_000231.
 SQ SEQUENCE 196 AA; 21959 MW; 8E9DE372A13669F4 CRC64;

Query Match 46.9%; Score 46; DB 1; Length 196;
 Best Local Similarity 50.0%; Pred. No. 5.3;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPFQVQS 18
 || ::| ||||| :
 Db 121 PPSAQPFQPFQQA 136

RESULT 3

Y258_HUMAN

ID Y258_HUMAN STANDARD; PRT; 391 AA.

AC Q92546;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Hypothetical protein KIAA0258.

GN KIAA0258.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.";
 RL DNA Res. 3:321-329(1996).

RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D87447; BAA13388.1; -.
 DR EMBL; BC001725; AAH01725.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 391 AA; 42455 MW; CE8F96D22A53D92A CRC64;

Query Match 45.9%; Score 45; DB 1; Length 391;
 Best Local Similarity 63.6%; Pred. No. 16;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLP 13
 ||| :||:|
 Db 64 PPDSSQPDVQP 74

RESULT 4

H105_CRIGR

ID H105_CRIGR STANDARD; PRT; 858 AA.
 AC Q60446;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Heat-shock protein 105 kDa (Heat shock 110 kDa protein).
 GN HSPH1 OR HSP105 OR HSP110.
 OS Cricetulus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95318163; PubMed=7797574;
 RA Lee-Yoon D., Easton D., Murawski M., Burd R., Subject J.R.;
 RT "Identification of a major subfamily of large hsp70-like proteins
 RT through the cloning of the mammalian 110-kDa heat shock protein.";
 RL J. Biol. Chem. 270:15725-15733(1995).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN, AND IS
 CC ALSO FOUND IN THE LIVER.
 CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z47807; CAA87768.1; -.
 DR PIR; A57513; A57513.
 DR InterPro; IPR001023; Hsp70.
 DR Pfam; PF00012; HSP70; 1.
 DR PRINTS; PR00301; HEATSHOCK70.
 DR ProDom; PD000089; Hsp70; 3.
 DR PROSITE; PS00297; HSP70_1; FALSE_NEG.
 DR PROSITE; PS00329; HSP70_2; FALSE_NEG.
 DR PROSITE; PS01036; HSP70_3; 1.
 KW ATP-binding; Heat shock; Multigene family.
 SQ SEQUENCE 858 AA; 96151 MW; 33B3CD01A97162DE CRC64;

Query Match 45.9%; Score 45; DB 1; Length 858;
 Best Local Similarity 41.2%; Pred. No. 36;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 DQPPDVEKPDLPFQVQ 17
 ||||: ||| :: |:
 Db 580 DQPPEAKKPKIKVVNVE 596

RESULT 5

H105_HUMAN

ID H105_HUMAN STANDARD; PRT; 858 AA.
 AC Q92598; O95739; Q9UPC4;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Heat-shock protein 105 kDa (Heat shock 110 kDa protein) (Antigen
 DE NY-CO-25).
 GN HSPH1 OR HSP105 OR HSP110 OR KIAA0201.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.";
 RL DNA Res. 3:321-329(1996).
 RN [2]
 RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
 RX MEDLINE=99132026; PubMed=9931472;
 RA Ishihara K., Yasuda K., Hatayama T.;
 RT "Molecular cloning, expression and localization of human 105 kDa heat
 RT shock protein, hsp105.";
 RL Biochim. Biophys. Acta 1444:138-142(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colorectal carcinoma;
 RX MEDLINE=98272252; PubMed=9610721;
 RA Scanlan M.J., Chen Y.-T., Williamson B., Gure A.O., Stockert E.,
 RA Gordan J.D., Tuereci O., Sahin U., Pfreundschuh M., Old L.J.;
 RT "Characterization of human colon cancer antigens recognized by
 RT autologous antibodies.";
 RL Int. J. Cancer 76:652-658(1998).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
 RC TISSUE=Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Alpha;
 CC IsoId=Q92598-1; Sequence=Displayed;
 CC Name=Beta;
 CC IsoId=Q92598-2; Sequence=VSP_002428;
 CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D86956; BAA13192.1; -.
 DR EMBL; AB003334; BAA34780.1; -.
 DR EMBL; AB003333; BAA34779.1; -.
 DR EMBL; AF039695; AAC18044.1; ALT_INIT.
 DR EMBL; BC037553; AAH37553.1; -.
 DR Genew; HGNC:16969; HSPH1.
 DR GO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0003773; F:heat shock protein activity; TAS.
 DR InterPro; IPR001023; Hsp70.
 DR Pfam; PF00012; HSP70; 1.
 DR PRINTS; PR00301; HEATSHOCK70.
 DR ProDom; PD000089; Hsp70; 3.
 DR PROSITE; PS00297; HSP70_1; FALSE_NEG.
 DR PROSITE; PS00329; HSP70_2; FALSE_NEG.
 DR PROSITE; PS01036; HSP70_3; 1.
 KW ATP-binding; Heat shock; Multigene family; Alternative splicing.
 FT VARSPLIC 529 572 Missing (in isoform Beta).
 FT /FTId=VSP_002428.
 SQ SEQUENCE 858 AA; 96864 MW; D0E757970E340B56 CRC64;

Query Match 45.9%; Score 45; DB 1; Length 858;
 Best Local Similarity 41.2%; Pred. No. 36;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFQVQ 17
 ||||: ||: |:
 Db 579 DQPPEAKKPKIKVVNVE 595

RESULT 6

H105_MOUSE

ID H105_MOUSE STANDARD; PRT; 858 AA.
 AC Q61699; Q62578; Q62579;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Heat-shock protein 105 kDa (Heat shock-related 100 kDa protein E7I)
 DE (HSP-E7I) (Heat shock 110 kDa protein) (42 degrees C-HSP).
 GN HSPH1 OR HSP105 OR HSP110.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96013135; PubMed=7556594;
 RA Morozov A., Subjeck J., Raychaudhuri P.;
 RT "HPV16 E7 oncoprotein induces expression of a 110 kDa heat shock
 RT protein.";
 RL FEBS Lett. 371:214-218(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96102018; PubMed=8530361;
 RA Yasuda K., Nakai A., Hatayama T., Nagata K.;
 RT "Cloning and expression of murine high molecular mass heat shock
 RT proteins, HSP105.";
 RL J. Biol. Chem. 270:29718-29723(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM HSP105-ALPHA).
 RC STRAIN=BALB/c;
 RX MEDLINE=99167340; PubMed=10066425;
 RA Yasuda K., Ishihara K., Nakashima K., Hatayama T.;
 RT "Genomic cloning and promoter analysis of the mouse 105-kDa heat shock
 RT protein (HSP105) gene.";
 RL Biochem. Biophys. Res. Commun. 256:75-80(1999).
 CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=HSP105-alpha;
 CC IsoId=Q61699-1; Sequence=Displayed;
 CC Name=HSP105-beta;
 CC IsoId=Q61699-2; Sequence=VSP_002429;
 CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES. HIGHLY EXPRESSED IN
 CC BRAIN.
 CC -!- INDUCTION: BY HEAT SHOCK. HSP105-ALPHA ALSO INDUCED BY OTHER
 CC STRESSES.
 CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L40406; AAA99485.1; -.
 DR EMBL; D67016; BAA11035.1; -.
 DR EMBL; D67017; BAA11036.1; -.
 DR EMBL; AB005282; BAA74540.1; -.
 DR EMBL; AB005267; BAA74540.1; JOINED.
 DR EMBL; AB005268; BAA74540.1; JOINED.

DR EMBL; AB005269; BAA74540.1; JOINED.
 DR EMBL; AB005270; BAA74540.1; JOINED.
 DR EMBL; AB005271; BAA74540.1; JOINED.
 DR EMBL; AB005272; BAA74540.1; JOINED.
 DR EMBL; AB005273; BAA74540.1; JOINED.
 DR EMBL; AB005274; BAA74540.1; JOINED.
 DR EMBL; AB005275; BAA74540.1; JOINED.
 DR EMBL; AB005276; BAA74540.1; JOINED.
 DR EMBL; AB005277; BAA74540.1; JOINED.
 DR EMBL; AB005278; BAA74540.1; JOINED.
 DR EMBL; AB005279; BAA74540.1; JOINED.
 DR EMBL; AB005280; BAA74540.1; JOINED.
 DR EMBL; AB005281; BAA74540.1; JOINED.
 DR PIR; S66666; S66666.
 DR MGD; MGI:105053; Hsp105.
 DR InterPro; IPR001023; Hsp70.
 DR Pfam; PF00012; HSP70; 1.
 DR PRINTS; PR00301; HEATSHOCK70.
 DR ProDom; PD000089; Hsp70; 3.
 DR PROSITE; PS00297; HSP70_1; FALSE_NEG.
 DR PROSITE; PS00329; HSP70_2; FALSE_NEG.
 DR PROSITE; PS01036; HSP70_3; 1.
 KW ATP-binding; Heat shock; Multigene family; Alternative splicing.
 FT VARSPLIC 530 573 Missing (in isoform HSP105-beta).
 FT /FTId=VSP_002429.
 FT CONFLICT 7 8 DV -> EL (IN REF. 1).
 FT CONFLICT 159 159 R -> A (IN REF. 2; BAA11035 AND 3;
 FT BAA74540).
 FT CONFLICT 320 320 P -> L (IN REF. 2; BAA11036).
 FT CONFLICT 373 373 A -> R (IN REF. 1).
 FT CONFLICT 518 518 P -> FQ (IN REF. 1).
 FT CONFLICT 744 744 I -> N (IN REF. 1).
 FT CONFLICT 838 838 A -> R (IN REF. 1).
 SQ SEQUENCE 858 AA; 96492 MW; 48D668236D8D3E17 CRC64;

Query Match 45.9%; Score 45; DB 1; Length 858;
 Best Local Similarity 41.2%; Pred. No. 36;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 DQPPDVEKPDLPFQVQ 17
 ||||: :|| :: |:
 Db 580 DQPPEAKKPKIKVVNVE 596

RESULT 7

DNL1_XENLA

ID DNL1_XENLA STANDARD; PRT; 1070 AA.

AC P51892;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DNA ligase I (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP]).

GN LIG1 OR LIGI.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96269417; PubMed=8682316;
 RA Lepetit D., Thiebaud P., Aoufouchi S., Prigent C., Guesne R.,
 RA Theze N.;
 RT "The cloning and characterization of a cDNA encoding *Xenopus laevis*
 RT DNA ligase I.";
 RL Gene 172:273-277(1996).
 CC -!- FUNCTION: This protein seals, during DNA replication, DNA
 CC recombination and DNA repair, nicks in double-stranded DNA.
 CC -!- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide}(N) +
 CC {deoxyribonucleotide}(M) = AMP + diphosphate +
 CC {deoxyribonucleotide}(N+M).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the ATP-dependent DNA ligase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L43496; AAB37754.1; -.
 DR PIR; JC4852; JC4852.
 DR InterPro; IPR000977; DNA_ligase.
 DR Pfam; PF01068; DNA_ligase; 1.
 DR Pfam; PF04679; DNA_ligase_A C; 1.
 DR Pfam; PF04675; DNA_ligase_A N; 1.
 DR TIGRFAMs; TIGR00574; dnll; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
 DR PROSITE; PS50160; DNA_LIGASE_A3; 1.
 KW Ligase; DNA repair; DNA replication; DNA recombination; Cell division;
 KW ATP-binding; Nuclear protein.
 FT BINDING 721 721 AMP (BY SIMILARITY).
 SQ SEQUENCE 1070 AA; 120233 MW; 065D6E5B8C6E4E52 CRC64;

Query Match 44.9%; Score 44; DB 1; Length 1070;
 Best Local Similarity 66.7%; Pred. No. 65;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 DVEKPDLPFQV 16
 | || :||||
 Db 779 DAEKKQIQPFQV 790

RESULT 8

SAH1_MOUSE

ID SAH1_MOUSE STANDARD; PRT; 1230 AA.
 AC P59808;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE SAM and SH3 domains containing protein 1.

GN SASH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22656757; PubMed=12771949;
 RA Zeller C., Hinzmann B., Seitz S., Prokoph H., Burkhardt-Goettges E.,
 RA Fischer J., Jandrig B., Estevez-Schwarz L., Rosenthal A.,
 RA Scherneck S.;
 RT "SASH1 - a candidate tumour suppressor gene on chromosome 6q24.3 is
 RT downregulated in breast cancer.";
 RL Oncogene 22:2972-2983(2003).
 CC -!- FUNCTION: May have a role in a signaling pathway. Could act as a
 CC tumor supressor.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -!- SIMILARITY: Contains 2 sterile alpha motif (SAM) domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ507736; CAD47812.1; -.
 DR InterPro; IPR001660; SAM.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00536; SAM; 1.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00454; SAM; 2.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50105; SAM_DOMAIN; 2.
 DR PROSITE; PS50002; SH3; FALSE_NEG.
 KW Anti-oncogene; SH3 domain; Repeat.
 FT DOMAIN 550 607 SH3 (BY SIMILARITY).
 FT DOMAIN 626 690 SAM 1.
 FT DOMAIN 972 1042 PRO-RICH.
 FT DOMAIN 1160 1224 SAM 2.
 SQ SEQUENCE 1230 AA; 135590 MW; DDE421DB74FE49AF CRC64;

Query Match 44.9%; Score 44; DB 1; Length 1230;
 Best Local Similarity 56.2%; Pred. No. 75;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 2 QPPDVEKPDLPFQVQ 17
 | ||||| | :|
 Db 84 QDL DVEKPDASPTSLQ 99

RESULT 9
 NP14_HUMAN
 ID NP14_HUMAN STANDARD; PRT; 699 AA.
 AC Q14978; Q15030;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
 DE nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
 DE phosphoprotein 1).
 GN NOLC1 OR KIAA0035.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
 RC TISSUE=Leukemia;
 RX MEDLINE=95386590; PubMed=7657714;
 RA Pai C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;
 RT "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar
 RT protein p130 are associated with nucleologenesis.";
 RL J. Cell Sci. 108:1911-1920(1995).
 RN [2]
 RP SEQUENCE OF 3-699 FROM N.A. (ISOFORM BETA).
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051387; PubMed=7584026;
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. I.
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
 RT analysis of randomly sampled cDNA clones from human immature myeloid
 RT cell line KG-1.";
 RL DNA Res. 1:27-35(1994).
 RN [3]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=96205319; PubMed=8630004;
 RA Pai C.-Y., Yeh N.-H.;
 RT "Cell proliferation-dependent expression of two isoforms of the
 RT nucleolar phosphoprotein p130.";
 RL Biochem. Biophys. Res. Commun. 221:581-587(1996).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=97168979; PubMed=9016786;
 RA Chen H.-K., Yeh N.-H.;
 RT "The nucleolar phosphoprotein P130 is a GTPase/ATPase with intrinsic
 RT property to form large complexes triggered by F⁻ and Mg²⁺.";
 RL Biochem. Biophys. Res. Commun. 230:370-375(1997).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=20036810; PubMed=10567578;
 RA Chen H.-K., Pai C.-Y., Huang J.-Y., Yeh N.-H.;
 RT "Human Nopp140, which interacts with RNA polymerase I: implications
 RT for rRNA gene transcription and nucleolar structural organization.";
 RL Mol. Cell. Biol. 19:8536-8546(1999).
 CC -!- FUNCTION: Related to nucleologenesis, may play a role in the
 CC maintenance of the fundamental structure of the fibrillar center
 CC and dense fibrillar component in the nucleolus. It has intrinsic
 CC GTPase and ATPase activities. May play an important role in
 CC transcription catalyzed by RNA polymerase I.
 CC -!- SUBUNIT: Interacts with RNA polymerase I 194 kDa subunit (RPA194)
 CC and with casein kinase-II.

```

CC  -!- SUBCELLULAR LOCATION: Shuttles between the nucleolus and the
CC      cytoplasm. At telophase it begins to assemble into granular-like
CC      pre-nucleolar bodies which are subsequently relocated to nucleoli
CC      at the early G1-phase.
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=Alpha;
CC      IsoId=Q14978-1; Sequence=Displayed;
CC      Name=Beta;
CC      IsoId=Q14978-2; Sequence=VSP_004338;
CC  -!- PTM: Undergoes rapid and massive phosphorylation/dephosphorylation
CC      cycles on CK2 and PKC sites. There is evidence suggesting that
CC      CDC2 kinase phosphorylates p130 at the M-phase.
CC  -!- SIMILARITY: Contains 1 Lish domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z34289; CAA84063.1; -.
DR  EMBL; D21262; BAA04803.1; -.
DR  PIR; I38073; I38073.
DR  Genew; HGNC:15608; NOLC1.
DR  GK; Q14978; -.
DR  MIM; 602394; -.
DR  GO; GO:0005737; C:cytoplasm; TAS.
DR  GO; GO:0005730; C:nucleolus; TAS.
DR  GO; GO:0007049; P:cell cycle; TAS.
DR  GO; GO:0007067; P:mitosis; TAS.
DR  GO; GO:0006364; P:rRNA processing; TAS.
DR  InterPro; IPR006594; Lish.
DR  InterPro; IPR007718; SRP40_C.
DR  Pfam; PF05022; SRP40_C; 1.
DR  SMART; SM00667; Lish; 1.
DR  PROSITE; PS50896; LISH; 1.
KW  Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;
KW  Alternative splicing.
FT  DOMAIN      10      42      LISH.
FT  DOMAIN      84     566     11 X 12 AA APPROXIMATE REPEATS OF AN
FT                                     ACIDIC SERINE CLUSTER.
FT  REPEAT      84      95      ACIDIC SERINE CLUSTER 1.
FT  REPEAT     125     136      ACIDIC SERINE CLUSTER 2.
FT  REPEAT     167     178      ACIDIC SERINE CLUSTER 3.
FT  REPEAT     221     232      ACIDIC SERINE CLUSTER 4.
FT  REPEAT     264     275      ACIDIC SERINE CLUSTER 5.
FT  REPEAT     325     336      ACIDIC SERINE CLUSTER 6.
FT  REPEAT     363     375      ACIDIC SERINE CLUSTER 7.
FT  REPEAT     425     436      ACIDIC SERINE CLUSTER 8.
FT  REPEAT     470     481      ACIDIC SERINE CLUSTER 9.
FT  REPEAT     519     529      ACIDIC SERINE CLUSTER 10.
FT  REPEAT     555     566      ACIDIC SERINE CLUSTER 11.
FT  DOMAIN      68      82      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  DOMAIN     204     382      INTERACTS WITH RPA194.

```

FT	DOMAIN	384	587	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT	DOMAIN	601	617	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT	MOD_RES	563	563	PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT	VARSPPLIC	241	241	K -> KVWTITSVRAE (in isoform Beta).
FT				/FTId=VSP_004338.
FT	CONFLICT	3	3	D -> A (IN REF. 2).
FT	CONFLICT	133	133	R -> S (IN REF. 2).
FT	CONFLICT	291	292	YA -> SV (IN REF. 2).
FT	CONFLICT	456	456	S -> P (IN REF. 2).
SQ	SEQUENCE	699 AA; 73720 MW; DFD4AD94EDF659FB CRC64;		

Query Match 43.9%; Score 43; DB 1; Length 699;
 Best Local Similarity 38.9%; Pred. No. 58;
 Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLQPFQVQS 18
 |:|| :|| :| |::
 Db 177 DEPPKNQKPKITPVTVKA 194

RESULT 10

Z407_HUMAN

ID Z407_HUMAN STANDARD; PRT; 1165 AA.
 AC Q9C0G0;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 407 (Fragment).
 GN ZNF407 OR KIAA1703.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=21082932; PubMed=11214970;
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:347-355(2000).
 CC -!- FUNCTION: May function as a transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: Contains 12 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB051490; BAB21794.1; -.
 DR HSSP; P08047; 1SP2.
 DR Genew; HGNC:19904; ZNF407.

DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 11.
 DR SMART; SM00355; Znf_C2H2; 12.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT NON_TER 1 1
 FT ZN_FING 331 354 C2H2-TYPE 1 (ATYPICAL).
 FT ZN_FING 361 385 C2H2-TYPE 2.
 FT ZN_FING 403 426 C2H2-TYPE 3.
 FT ZN_FING 454 478 C2H2-TYPE 4 (ATYPICAL).
 FT ZN_FING 484 506 C2H2-TYPE 5.
 FT ZN_FING 512 535 C2H2-TYPE 6.
 FT ZN_FING 545 567 C2H2-TYPE 7.
 FT ZN_FING 573 597 C2H2-TYPE 8.
 FT ZN_FING 603 625 C2H2-TYPE 9.
 FT ZN_FING 631 653 C2H2-TYPE 10.
 FT ZN_FING 659 684 C2H2-TYPE 11 (ATYPICAL).
 FT ZN_FING 690 713 C2H2-TYPE 12 (ATYPICAL).
 SQ SEQUENCE 1165 AA; 126980 MW; A37B8A9701F5133E CRC64;

Query Match 43.9%; Score 43; DB 1; Length 1165;
 Best Local Similarity 63.6%; Pred. No. 1e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDL 11
 :| ||:| |||
 Db 711 EQHPDIENPDL 721

RESULT 11

DPOA_HUMAN

ID DPOA_HUMAN STANDARD; PRT; 1462 AA.
 AC P09884;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE DNA polymerase alpha catalytic subunit (EC 2.7.7.7).
 GN POLA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88196090; PubMed=3359994;
 RA Wong S.W., Wahl A.F., Yuan P.-M., Arai N., Pearson B.E., Arai K.,
 RA Korn D., Hunkapiller M.W., Wang T.S.-F.;
 RT "Human DNA polymerase alpha gene expression is cell proliferation
 RT dependent and its primary structure is similar to both prokaryotic
 RT and eukaryotic replicative DNA polymerases."
 RL EMBO J. 7:37-47(1988).
 RN [2]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=91172197; PubMed=2005899;
 RA Pearson B.E., Nasheuer H.P., Wang T.S.;

```

RT   "Human DNA polymerase alpha gene: sequences controlling expression in
RT   cycling and serum-stimulated cells.";
RL   Mol. Cell. Biol. 11:2081-2095(1991).
CC   -!- FUNCTION: Polymerase alpha in a complex with DNA primase is a
CC       replicative polymerase.
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
CC       alpha, beta, gamma, delta, and epsilon which are responsible for
CC       different reactions of DNA synthesis.
CC   -!- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X06745; CAA29920.1; -.
DR   EMBL; M64481; AAA52318.1; -.
DR   PIR; S00257; DJHUAC.
DR   Genew; HGNC:9173; POLA.
DR   GK; P09884; -.
DR   MIM; 312040; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003889; F:alpha DNA polymerase activity; NAS.
DR   GO; GO:0006260; P:DNA replication; NAS.
DR   InterPro; IPR006172; DNA_pol_B.
DR   InterPro; IPR006134; DNA_pol_B_dom.
DR   InterPro; IPR006133; DNA_pol_B_exo.
DR   InterPro; IPR004578; Pol2.
DR   Pfam; PF00136; DNA_pol_B; 1.
DR   Pfam; PF03104; DNA_pol_B_exo; 1.
DR   PRINTS; PR00106; DNAPOLB.
DR   SMART; SM00486; POLBc; 1.
DR   TIGRFAMs; TIGR00592; pol2; 1.
DR   PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   DNA-binding; Nuclear protein.
FT   DNA_BIND    650    715    POTENTIAL.
FT   DNA_BIND    1245   1376    POTENTIAL.
SQ   SEQUENCE    1462 AA;  165860 MW;  25C270B0A0DB38BE CRC64;

```

```

Query Match          43.9%;  Score 43;  DB 1;  Length 1462;
Best Local Similarity 38.9%;  Pred. No. 1.3e+02;
Matches      7;  Conservative      6;  Mismatches      5;  Indels      0;  Gaps      0;

```

```

QY          1 DQPPDVEKPDLPFQVQS 18
              |:| :||: ||:|  ::
Db          255 DEPMEVEEVDLEPMAAKA 272

```

RESULT 12
MAPA_RAT

ID MAPA_RAT STANDARD; PRT; 2774 AA.
 AC P34926;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1A (MAP 1A) [Contains: MAP1 light chain
 DE LC2].
 GN MAP1A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92355629; PubMed=1379599;
 RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
 RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
 RT one messenger RNA.";
 RL J. Biol. Chem. 267:16561-16566(1992).
 CC -!- FUNCTION: Structural protein involved in the filamentous cross-
 CC bridging between microtubules and other skeletal elements.
 CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
 CC with MAP1A and MAP1B proteins.
 CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
 CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
 CC THEIR MORPHOLOGY.
 CC -!- DOMAIN: The basic region containing the repeats may be responsible
 CC for the binding of MAP1A to microtubules.
 CC -!- PTM: Various serine residues may be phosphorylated by cAMP kinase.
 CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B.
 CC -!- SIMILARITY: TO MAP1B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M83196; AAB48069.1; -.
 DR PIR; A43359; A43359.
 KW Microtubule; Repeat; Phosphorylation.
 FT CHAIN ?2465 2774 MAP1 LIGHT CHAIN LC2.
 FT DOMAIN 309 496 LYS-RICH (BASIC).
 FT DOMAIN 336 541 11 X 3 AA REPEATS OF K-K-[DE].
 FT REPEAT 336 338 1.
 FT REPEAT 415 417 2.
 FT REPEAT 420 422 3.
 FT REPEAT 424 426 4.
 FT REPEAT 427 429 5.
 FT REPEAT 431 433 6.
 FT REPEAT 436 438 7.

FT REPEAT 440 442 8.
 FT REPEAT 444 446 9.
 FT REPEAT 449 451 10.
 FT REPEAT 539 541 11.
 SQ SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 43.4%; Score 42.5; DB 1; Length 2774;
 Best Local Similarity 52.9%; Pred. No. 3e+02;
 Matches 9; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

Qy 1 DQPPDVE---KPDLPQPF 14
 | |::: |||:|
 Db 451 DTKPEVKKLSKPDLPKF 467

RESULT 13

SLBP_XENLA

ID SLBP_XENLA STANDARD; PRT; 254 AA.

AC P79943;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Histone RNA hairpin-binding protein (Histone stem-loop binding protein 1).

GN SLBP1 OR SLBP OR HBP.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Oocyte;

RX MEDLINE=97115884; PubMed=8957003;

RA Wang Z.-F., Whitfield M.L., Ingledue T.C. III, Dominski Z.,

RA Marzluff W.F.;

RT "The protein that binds the 3' end of histone mRNA: a novel RNA-

RT binding protein required for histone pre-mRNA processing.";

RL Genes Dev. 10:3028-3040(1996).

RN [2]

RP PHOSPHORYLATION.

RX MEDLINE=20387311; PubMed=10827192;

RA Mueller B., Link J., Smythe C.;

RT "Assembly of U7 small nuclear ribonucleoprotein particle and histone

RT RNA 3' processing in Xenopus egg extracts.";

RL J. Biol. Chem. 275:24284-24293(2000).

CC -!- FUNCTION: BINDS THE STEM-LOOP STRUCTURE OF REPLICATION-DEPENDENT
 CC HISTONE PRE-MRNAs AND CONTRIBUTES TO EFFICIENT 3' END PROCESSING
 CC BY STABILIZING THE COMPLEX BETWEEN HISTONE PRE-MRNA AND U7 SMALL
 CC NUCLEAR RIBONUCLEOPROTEIN (SNRNP) (BY SIMILARITY). COULD PLAY AN
 CC IMPORTANT ROLE IN TARGETING MATURE HISTONE MRNA FROM THE NUCLEUS
 CC TO THE CYTOPLASM AND TO THE TRANSLATION MACHINERY. STABILIZES
 CC MATURE HISTONE MRNA AND COULD BE INVOLVED IN CELL-CYCLE REGULATION
 CC OF HISTONE GENE EXPRESSION.

CC -!- SUBCELLULAR LOCATION: NUCLEAR (COILED BODIES) AND CYTOPLASMIC.

CC -!- TISSUE SPECIFICITY: Widely expressed.

CC -!- DEVELOPMENTAL STAGE: VERY LOW LEVELS IN STAGE I OOCYTES, GRADUALLY

CC INCREASING THROUGHOUT OOGENESIS. FURTHER INCREASE IS ACHIEVED
CC DURING EARLY EMBRYOGENESIS.
CC -!- PTM: Phosphorylated on Thr-60 during mitosis.
CC -!- SIMILARITY: BELONGS TO THE SLBP FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U75681; AAC60342.1; -.
KW RNA-binding; mRNA processing; Nuclear protein; Phosphorylation.
FT MOD_RES 60 60 PHOSPHORYLATION (BY CDC2).
FT DOMAIN 127 196 RNA-BINDING (BY SIMILARITY).
SQ SEQUENCE 254 AA; 29726 MW; DFA0651D13D55B0C CRC64;

Query Match 42.9%; Score 42; DB 1; Length 254;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PPDVEKPDLPQPF 14
|| | ||||
Db 197 PPAAEGSDLPQPF 208

RESULT 14

IG1R_PIG

ID IG1R_PIG STANDARD; PRT; 304 AA.
AC Q29000; Q28951;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Insulin-like growth factor I receptor (EC 2.7.1.112) (Fragments).
GN IGF1R.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE OF 1-186 FROM N.A.
RC TISSUE=Skeletal muscle;
RA Matteri R.L., Anderson J.E., Prather R.S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 187-304 FROM N.A.
RC TISSUE=Conceptus membrane;
RX MEDLINE=95377227; PubMed=7649105;
RA Green M.L., Simmen R.C.M., Simmen F.A.;
RT "Developmental regulation of steroidogenic enzyme gene expression in
RT the periimplantation porcine conceptus: a paracrine role for
RT insulin-like growth factor-I."
RL Endocrinology 136:3961-3970(1995).
CC -!- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)
CC WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A

```

CC      TYROSINE-PROTEIN KINASE ACTIVITY.
CC      -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC          tyrosine phosphate.
CC      -!- SUBUNIT: TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY DISULFIDE
CC          BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF THE LIGAND-
CC          BINDING DOMAIN, WHILE THE BETA CHAIN CARRIES THE KINASE DOMAIN.
CC      -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -!- SIMILARITY: Belongs to the Tyr family of protein kinases. Insulin
CC          receptor subfamily.
CC      -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U58370; AAB02578.1; -.
DR      EMBL; U15445; AAB49731.1; -.
DR      InterPro; IPR008957; FN_III-like.
DR      InterPro; IPR003961; FN_III.
DR      InterPro; IPR000719; Prot_kinase.
DR      InterPro; IPR002011; RecepttyrkinsII.
DR      InterPro; IPR008266; Tyr_pkinase_AS.
DR      ProDom; PD000001; Prot_kinase; 1.
DR      SMART; SM00060; FN3; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
DR      PROSITE; PS00239; RECEPTOR_TYR_KIN_II; PARTIAL.
KW      Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW      Glycoprotein; ATP-binding; Phosphorylation; Repeat.
FT      NON_TER      1      1
FT      DOMAIN      <1      147      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM      148      168      POTENTIAL.
FT      DOMAIN      169      >304      CYTOPLASMIC (POTENTIAL).
FT      CARBOHYD      115      115      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD      128      128      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      NON_CONS      186      187
FT      NON_TER      304      304
SQ      SEQUENCE      304 AA; 34615 MW; E026E01215FC3AB8 CRC64;

Query Match      42.9%; Score 42; DB 1; Length 304;
Best Local Similarity 38.9%; Pred. No. 34;
Matches      7; Conservative      7; Mismatches      4; Indels      0; Gaps      0;

Qy      1 DQPPDVEKPDLPFQVQS 18
      ::||: |: ||:| ::|
Db      232 NKPPEPEELDLEPENMES 249

RESULT 15
HXA2_MOUSE
ID      HXA2_MOUSE      STANDARD;      PRT;      372 AA.
AC      P31245;

```

DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-A2 (Hox-1.11).
 GN HOXA2 OR HOXA-2 OR HOX-1.11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gridley T.;
 RL Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92335281; PubMed=1352886;
 RA Tan D.P., Ferrante J., Nazarali A., Shao X., Kozak C.A., Guo V.,
 RA Nirenberg M.;
 RT "Murine Hox-1.11 homeobox gene structure and expression.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6280-6284(1992).
 RN [3]
 RP SEQUENCE OF 128-231 FROM N.A.
 RX MEDLINE=92212934; PubMed=1348361;
 RA Nazarali A., Kim Y., Nirenberg M.;
 RT "Hox-1.11 and Hox-4.9 homeobox genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:2883-2887(1992).
 CC -!- FUNCTION: Sequence-specific transcription factor which is part of
 CC a developmental regulatory system that provides cells with
 CC specific positional identities on the anterior-posterior axis.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.
 CC PROBOSCIPEDIA SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M95599; AAA37827.1; -.
 DR EMBL; M93148; AAA37835.1; -.
 DR EMBL; M93292; AAA37836.1; -.
 DR EMBL; M87801; AAA37834.1; -.
 DR PIR; A46037; A46037.
 DR HSSP; P14653; 1B72.
 DR TRANSFAC; T01698; -.
 DR MGD; MGI:96174; Hoxa2.
 DR InterPro; IPR001827; Antennapedia.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000047; HTH_lambrepressr.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00025; ANTENNAPEDIA.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 DR PROSITE; PS00032; ANTENNAPEDIA; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation.
 FT SITE 96 101 ANTP-TYPE HEXAPEPTIDE.
 FT DNA_BIND 139 198 HOMEBOX.
 SQ SEQUENCE 372 AA; 40793 MW; 0ADA79113DB72726 CRC64;

Query Match 42.9%; Score 42; DB 1; Length 372;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 1 DQPPDVEKPDLPFFQVQS 18
 | | : | | | | | |
 Db 308 DSPEALEVPSLQDFNVFS 325

Search completed: August 24, 2004, 15:43:25
 Job time : 11.6716 secs